Abdominal Aortic Aneurysm Study:

Creation and Validation of Chemical Imaging Algorithm

Honors Research Thesis

Presented in Partial Fulfillment of the Requirements for graduation

"with Honors Research Distinction in Chemistry" in the undergraduate colleges of The Ohio State University

by

Alexandra Lawson

The Ohio State University

May 2013

Project Advisor: J. Clay Harris, Noel M. Paul, Department of Chemistry and Biochemistry

Table of Contents

List of Figures

List of Tables

List of Equations

List of Schemes

Abstract:

Abdominal aortic aneurysms (AAAs) are known as a dilation of the abdominal aorta and often have no symptoms; due to this lack of symptoms, many go undiagnosed and as a result, approximately 15,000 Americans die each year due to AAA rupture. In order to develop a better method of detection, mice aorta samples were collected and analyzed with FT-IR microspectroscopy; the aorta samples included healthy aortas as well as induced AAAs. A chemical algorithm was written using MATLAB so that the data may be most efficiently analyzed, and a chemical difference could be determined between the healthy and diseased tissue. However, before the chemical algorithm is able to analyze the data, it first needed to be validated. Amino acid copolymers were synthesized as biochemical standards to test the algorithm. With the initial biochemical standards available, the algorithm has proved to be functioning properly, although further tests are needed. Determining a chemical difference between the healthy and diseased tissue will lead to the ultimate goal of developing a non-invasive tool for AAA detection.

Introduction:

Abdominal aortic aneurysms (AAAs) are often difficult to detect due to a lack of symptoms, and as a result, 15,000 people die each year of an AAA rupture in the United States alone.[1] An aneurysm "can be defined as a permanent and irreversible localized dilatation of a vessel"; an enlarged abdominal aorta is deemed an aneurysm once it has reached 1.5 times its normal diameter.[2] As shown in Figure 1, a true aneurysm involves all three layers of the aortic wall: adventitia, media and intima.[3] However, it is possible for only the adventitia layer to expand; it is considered a pseudoanuerysm, or false aneurysm, and it creates a separate pathway allowing blood flow.[3]

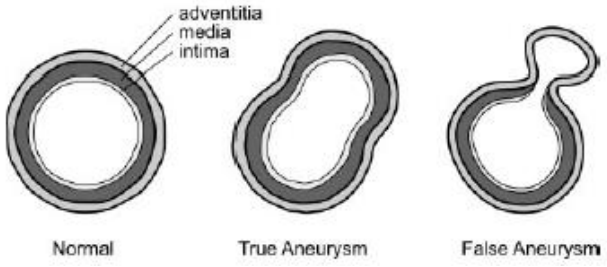


Figure 1. Types of aortic aneurysms.[3]

AAAs are often difficult to detect because they have little to no symptoms, and as a result, many individuals are not aware of their impending aneurysm or its possibility of rupture.[1] As the aneurysm grows, it causes portions of the arterial wall to thicken and other portions to weaken, as well as a shift in the collagen-elastin ratio.[2] There are two types of collagen that maintain the strength and structure of the arterial wall: collagen I and collagen III; the viscoelastic properties of the aorta are maintained by elastin and associated proteins, together forming elastic fibers.[2]

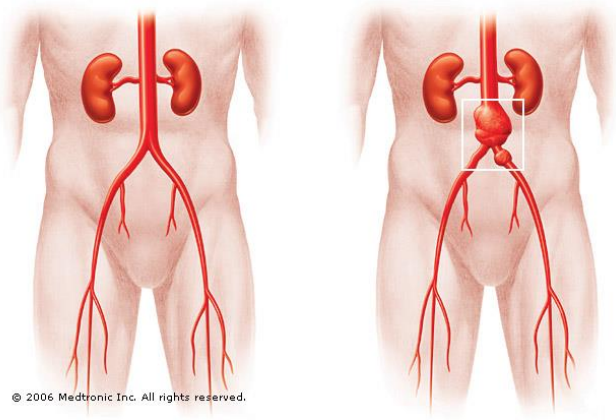


Figure 2. Normal aorta (left) and AAA (right).[4]

A shift in the collagen-elastin ratio changes the nature of the arterial wall, exposing some areas to be weakened or thin. With enough weakening, the wall will rupture, in which case medical attention must be sought immediately. "The overall mortality rate for patients with ruptured abdominal aortic aneurysms is between 65% and 85%, and about half of deaths attributed to rupture occur before the patient reaches the surgical room".[2]

Currently the direct causes of AAAs are unknown; however, there are several risk factors that have been linked to the development of AAAs. The most prominent risk factor is tobacco use; in fact, "the prevalence of abdominal aortic aneurysms in tobacco smokers is more than four times that in life-long nonsmokers".[2] Additional risk factors include a person's age and gender.[2] "The average age at the time of diagnosis is 65-70 years, and more men than women are affected".[3] Other risk factors include hypertension, hyperlipidaemia, chronic obstructive pulmonary disease and a family history of AAAs.[2]

These risk factors may be indicative of an AAA, but it can be difficult to diagnose because it has very few to no symptoms. An AAA can be diagnosed using a stethoscope and listening for the sounds of turbulent blood flow in the abdominal aorta; however this method is not always conclusive, and it is less effective in patients that are overweight.[1] Detection of an

AAA is mostly commonly achieved using imaging techniques such as an ultrasound, computed tomography (CT) scan, or magnetic resonance imaging (MRI).[3] CT scans are preferred because they are widely available and the images can be collected and interpreted within minutes.[3]

Once the AAA has been detected, there are several avenues for treatment. If the AAA is detected early in its development, physicians will encourage lifestyle changes, most importantly being cessation of tobacco use, in order to prevent further growth and rupture. Patients may also be prescribed non-steroidal anti-inflammatory drugs, statins, and pharmacologic inhibitors of matrix metalloproteinases (MMP), as these degrade collagen and elastic fibers.[2] A recent study has also shown that Angiotensin II inhibitors have been effective in slowing AAA progression.[5] Lastly, a physician will recommend surgery; however, this greatly depends on the stage of the AAA and the amount of risk factors because surgery poses a great risk to the patient.

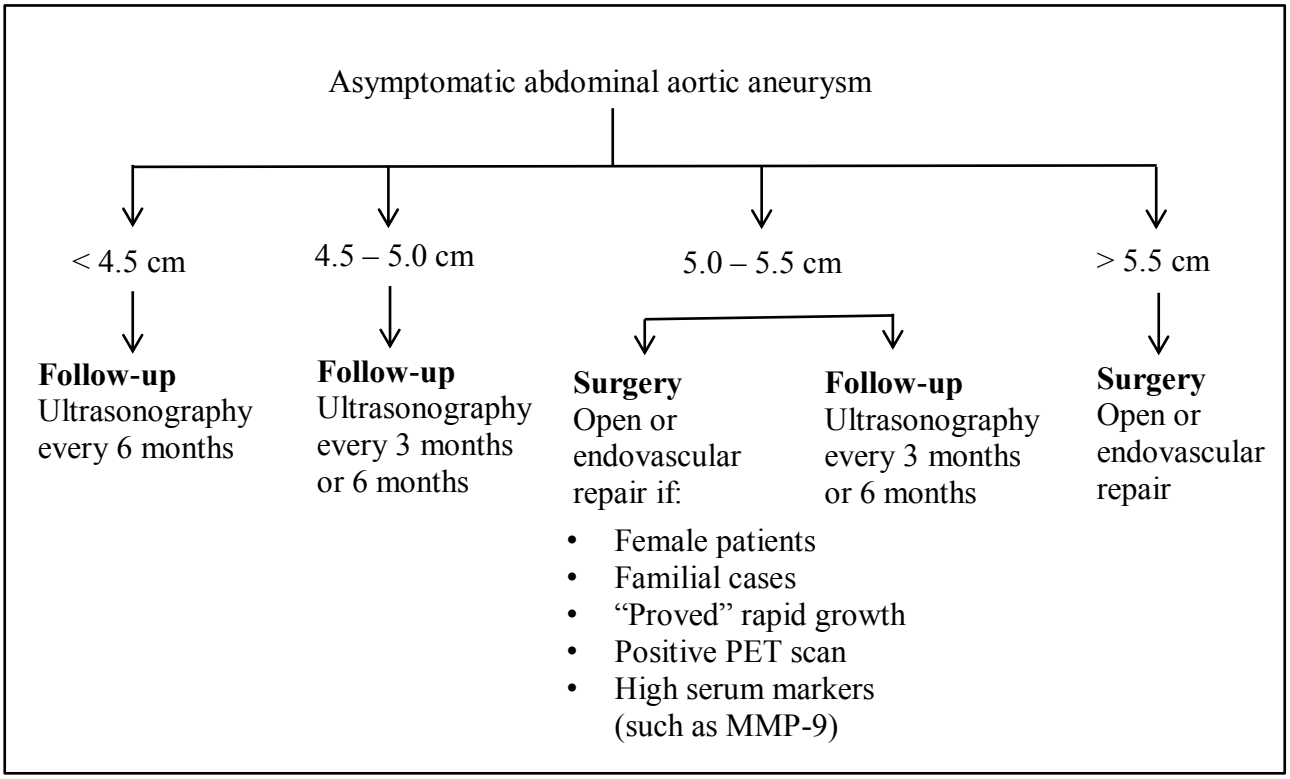


Figure 3. Management of an asymptomatic abdominal aortic aneurysm.[2]

Elective aneurysm repair has greatly improved in recent years. Essentially this current form of repair includes placing an endograft around the aneurysm. While it can vary depending on the hospital and surgeon, the "mean 30-day mortality rate has been reported at between 1.1% and 7.0%".[2] There is also a new treatment being researched that has signs of being safer and just as effective as conventional surgery.[6] A stent-graft device is placed in the area of the aneurysm though the use of a catheter.[6] The stent-graft device is attached to the catheter which enters the aorta through the femoral artery. Once the catheter reaches the aneurysm, the stent-graft design is released, causing it to bind to the walls of the aorta. It has shown promising results including less blood loss, lower recovery time and lower morbidity than conventional surgery.[6]

However, the key detail to elective aneurysm repairs, whether surgery or stent-graft inserted by catheter, is the awareness of the aneurysm to the individual. As mentioned before, there are very few to no symptoms associated with AAAs. Symptoms can present themselves with rupture of the aneurysm, but those symptoms "frequently are confused with those of appendicitis, pancreatitis, or bowel obstruction".[3] If the individual is able to seek immediate medical assistance, the outcome is still uncertain, as the mortality rate of emergency repair of a ruptured AAA is between 30-70%.[2] And "unlike coronary heart disease, the incidence of AAA is reported to be increasing in Western Countries".[5]

Currently, screening for AAA occurs primarily in individuals that have a family history of the disease. However, only 1-5% of patients with AAA have a family history of the disease.[5] As there are no direct connections to AAAs and other health conditions, only risk factors, it would be quite impractical to screen every person who demonstrates a risk factor of an AAA. Therefore, it is crucial to create a better method of AAA detection. In order to improve current

methods, it would ideal to create a method that is fairly noninvasive, but more importantly, can become a routine procedure.

Preliminary Studies and Work:

There have been several studies that explored the various causes and chemical compositions of AAAs. For instance, a study conducted in 1995 used laser-induced florescence (LF) to examine the atherosclerotic plaque composition in aorta tissue samples. Dogs were injected with several compounds including collagen I, III and IV, elastin and cholesterol; it was after the injection of the listed compounds, the researchers observed a great change in the LF intensity. This study demonstrated that LF is a valuable method for determining chemical composition of plaque, but it also indicated that these compounds are in some way involved with plaque accumulation in both the thoracic and abdominal aorta.[7]

A separate study focused on the chemical difference of inflammatory abdominal aortic aneurysms (IAAAs) and AAAs by analyzing tissue samples through light microscopy, immunohistochemistry, and transmission electron microscopy (TEM). Through these methods, it was determined that a large quantity of B lymphocytes were present in an IAAA as well as deposits of IgG, IgM, and C3c in the interstitial matrix. Lastly, there was "an increase in type III collagen and a reduction in elastin," suggesting the IAAA has a different development process than a typical AAA.[8]

In order to determine a chemical difference between a healthy aorta and aneurysm, aortic tissue samples were collected from mice and analyzed using Fourier Transform Infrared (FT-IR) microspectroscopy, prior to the author's contribution on this project.[1] "This study used apolipoprotein E (apoE) knockout mice ranging in age from 2 – 12 months".[9] The experimental group of mice were injected with angiotensin II to induce the aneurysm, a common procedure in cardiovascular research of animals.[9, 10]

[1] "All procedures involving animals were approved by the Animal Care and Use Committee at the University of Kentucky".[9]

> Two mm pieces from above (ak) and from below (bk) the kidney branch of the aorta from both an experimental mouse (x) and a control mouse (c) were frozen in OCT (Optimal Cutting Temperature compound). The 2 mm pieces were sequentially sectioned from top to bottom, taking 12 consecutive 5 μm slices from above and below the kidney in each mouse, resulting in 48 total samples (x-ak-1 to 12, x-bk-1 to 12, c-ak-1 to 12, and c-bk-1 to 12). These samples were mounted on two low-e glass slides (SensIR) for 2-D FTIR reflectance spectrometry.[9]

FT-IR microspectroscopy was chosen as the method of analysis because it a common yet effective method for identification of chemical composition. The common procedure for IR spectroscopy involves a sample is placed onto a salt plate, most common being sodium chloride or potassium bromide, and infrared light shines through the sample. When molecules are bombarded with light, they absorb some of the energy of that wavelength of light and express that energy. In the case of infrared light, the bonds between the atoms of molecules will increase their vibrational frequencies. Different types of bonds absorb different wavelengths of light and thus vibrate at different frequencies; these vibrations are recorded by the spectrometer, and a bond can be identified by the frequency and intensity of that vibration, ultimately leading to the identification of the molecule. The recorded vibrations are represented as a spectrum plotted as absorption against energy, which is the form of the data that will be analyzed in this study.

While this method is useful for qualitative analysis of the aortic tissue, it collects a substantial amount of data, as further explained in the next section. In order to competently analyze this data, an efficient method needed to be created, which is where the author's work begins.

Chemical Algorithm:

FT-IR microspectroscopy yields both special and spectra data. These datasets are often very large, creating significant challenges in analyzing experimental results. Figure 4 shows the image for one of the control mice above kidney cross-sections; this image is unique in that it is two-dimensional when the dataset is actually three-dimensional. The image shows the dimensions correspond to the length and width of the actual cross-section; the remaining dimension is the absorbance spectra from FT-IR microspectroscopy. In Figure 4, the absorbances of all energies in a specific pixel has been summed to give areas of high absorbance (blue) and areas of low absorbance (red), essentially showing the outline of the cross-sections.

The image is divided into what can be called pixels, similar to a normal photograph. However, instead of the pixel being a mixture of three colors like that of a photograph, the pixel is composed of an infrared spectrum. The values along the axes are pixel counts, therefore there are 6600 spectra for only the cdataak1 dataset. Figure 5, also a total absorbance image, shows that the xdataak1 dataset has over 12,000 spectra to be analyzed due to the experimental mice aortic cross-section being much larger than the control mice. With this amount of data, a better, more efficient method of analysis needed to be developed.

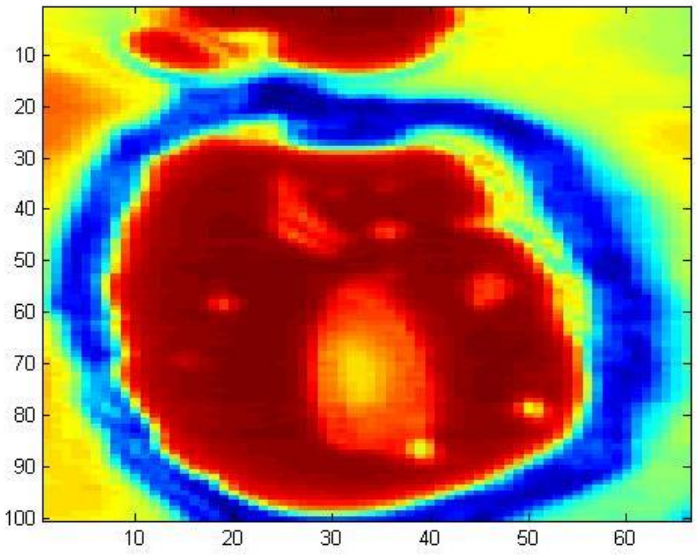


Figure 4. Total absorbance image of control data above kidney 1 (cdataak1).

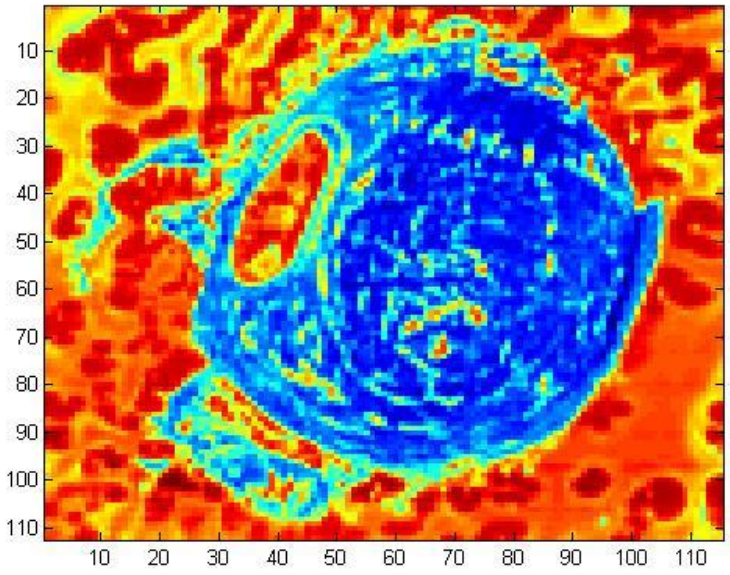


Figure 5. Total absorbance image of experiment data above kidney 1 (xdataak1).

Prior to writing the algorithm, the total absorbance images were created by taking the sum of the third (spectral) dimension. The data was originally not in this configuration when it was exported from the instrument. If the data were in the shape of a cube, the x and y dimension would be the length and width, respectfully, and the z dimension would be the amount of

wavenumbers used in the infrared light spectrum; for these samples, and spectrum was taken at every eighth wavenumber, creating a total of 411 wavenumbers in the z dimension. However, data received from the instrument was two dimensional: (x · y) by z. For example, the data received from cdataak1 was 6600 by 411. The instrument reports data in this way because it takes the spectra down a row, not necessarily all at once, and move to the next column after the row is completed.

As a result, the data of each x row is placed adjacent to each other into one long line of data, and the absorbance values fill the second dimension. In order to convert the data into a three-dimensional dataset, a loop first needed to be written that could manipulate the data into this form. Once it was in this three dimensional form, the total absorbance images were created. As the author was a novice to MATLAB® when starting this project, this beginning stage allowed adequate time and use to learn MATLAB® to a comfortable degree. After this stage, the chemical algorithm could be written.

The algorithm was written using MATLAB®, and due to the amount of data, it was determined that principal component analysis (PCA) and Mahalanobis distances would be the best methods to implement in the chemical algorithm for analyzing this aortic data. This combination of PCA and Mahalanobis distances has had previous success for analysis of gamma-ray flux car emissions; the instrument in question, a radiation portal monitor (RPM) and is used to first detect illicit nuclear material through its gamma-ray flux emissions and second, determine the location of that material in/on the vehicle. However, there are also benign gamma-ray emissions from vehicles that can trigger the RPM. The researchers collected signatures of benign gamma-ray emissions and used these signatures for comparison in a screening. By using

PCA and Mahalanobis distances, the RPM is able to determine which gamma-ray emissions are benign and which should cause alert, making the monitor useful.[11]

PCA is a method to reduce or simplify a dataset by "finding relationships between objects".[12] In other words, PCA will compare the data within a system to one another, and based on the amount of variance, rank them in order of greatest to least variance. Based on this ranking, PCA will output a new dataset for the original set of objects. The vectors of this matrix are called score vectors and loading vectors, and together these make a PC. The PC is then used as a dimension. Figure 6 shows an arbitrary plot of the first PC plotted against the second PC.

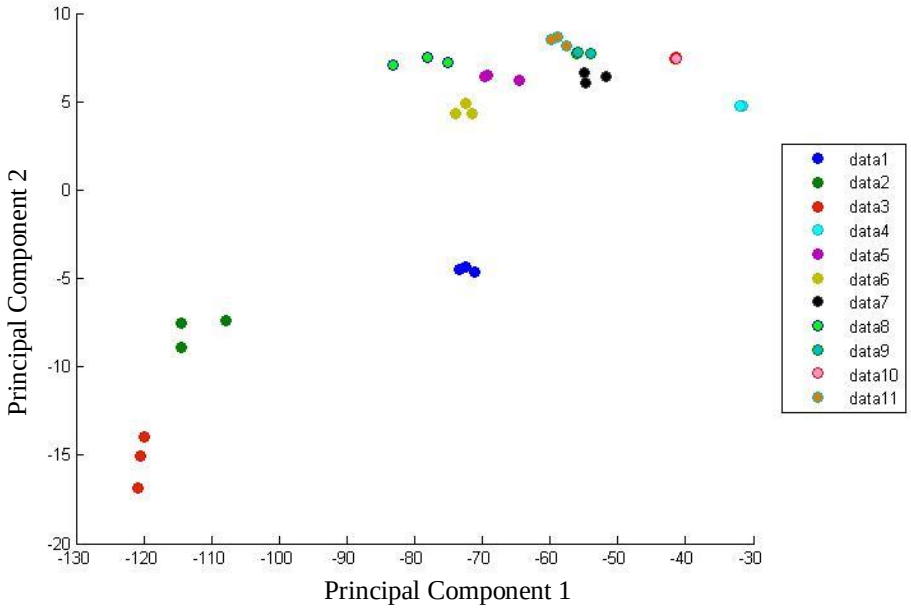


Figure 6. Example of generated PC comparison plot.

For the purpose of this example, the individual points can be considered IR spectra, with each point being a separate spectrum; the points are therefore color-coded based on the tissue sample. The data (or samples) that are most similar to each other will be plotted closer together, forming a cluster. In order to quantitatively measure the difference between these points, a Mahalanobis distance is used.

The Mahalanobis distance formula was considered along with the Euclidean distance formula for measuring distance between points; however, it was determined to be the better formula, as explained below. A specific spectrum will be chosen, which will be from here on referred to as the point of interest (POI). It will be used as the reference point from which all other points will be measured. In this sense, if the POI is the spectrum for molecule X, then the algorithm will measure the distances from the POI to the remaining points and determine which are chemically most similar and different to molecule X. When the algorithm has computed the PCA and Mahalanobis distances, it will return the results in a false color image based upon the calculated value; the color spectrum is shown in Figure 7. The dark blue end of the spectrum originally represented little to no difference to the POI, and dark red represented significant chemical difference. However, the gradient was later modified so that the reference is centered. It is important to identify negative and positive distances from the standard so the original false color gradient was modified. More detailed the reasoning is explained in the Creation of Biochemical Standard section.

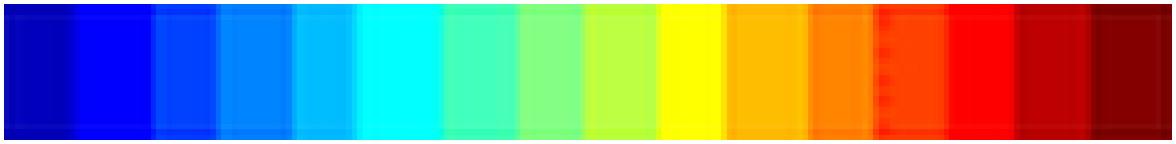


Figure 7. False-color gradient of Mahalanobis distances.

Euclidean distances (Equation 1) were also considered as a method of measurement, but there is one important piece missing from its formula. As shown below, the Mahalanobis distance formula (Equation 2) incorporates the standard deviation into the formula, so it returns the values in relation to a standard deviation.

$$d(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$$

Equation 1. Euclidean distance.

$$d(x,y) = \sqrt{\sum_{i=1}^{n}\frac{(x_i - y_i)^2}{\sigma_i^2}}$$

Equation 2. Mahalanobis distance.

Therefore, the Mahalanobis distance formula is capable of more accurately measuring a cluster that is not a perfect circle. Using Figure 8 as an example, the Euclidean distance formula would result in the vertical barriers to be a distance x, and the horizontal barriers to be a distance 2x. However, both the horizontal and vertical barriers are considered within the first standard deviation using the Mahalanobis distance formula.

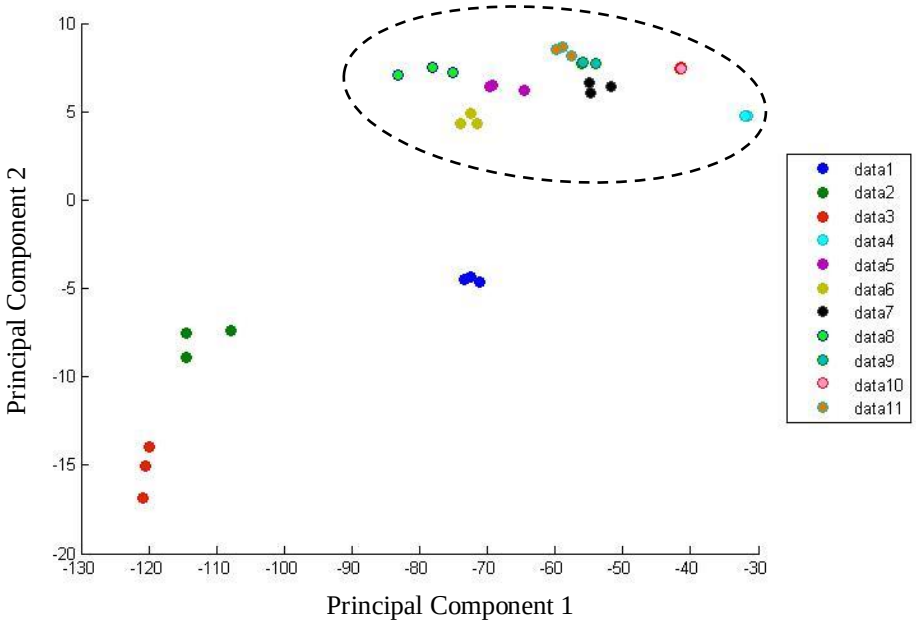


Figure 8. Example of generated PC comparison plot.

While this program is able to measure the statistical chemical difference between the different spectra, it is useless unless the spectrum of the POI is a known component. For

instance, the amount and location of collagen I can be determined in an aortic cross-section sample if the spectrum for collagen I is inserted as the reference or POI. It will return an image similar to that in Figures 4 and 5, but the false color gradient will represent the statistical difference from the collagen I.

The false color gradient provides a solution to an interesting problem with this algorithm. As previously said, the PCA program returns a new dataset in the form of PCs, with the variance within the dataset ranked from greatest to least variance. The majority of variances will be provided in the first few PCs. Therefore, the best comparison of the data would include more than the first two PCs; however, there is currently no way to visually represent a fourth dimension on a two-dimensional surface. The false color gradient is able to incorporate the fourth dimension and beyond into its comparison; those PCs need only be selected when running the chemical algorithm. Lastly, the algorithm also uses the MATLAB® functions reshape and imagesc to organize the Mahalanobis distance values and create the false-color image, respectfully.

Creation of Biochemical Standard:

Despite previously documented success with using PCA and Mahalanobis distances in analyzing large datasets, the chemical algorithm needed to be validated prior to analyzing the aortic tissue data. It was determined in order to most effectively test the algorithm, it would be best to use a material similar to the material that will ultimately be analyzed: the aortic tissue. As collagen and elastin are both known major components of the arterial wall, the building blocks, amino acids, were used for creating a biochemical standard. The process for creating the standard started by using four amino acids: glycine, L-proline, L-leucine, and L-valine. Glycine and L-proline were chosen because collagen generally consists of one-third glycine and one-sixth L-proline. L-leucine and L-valine were chosen because the IR spectra for each were quite similar (Figures 9 and 10); by using amino acids with similar IR spectra, it allowed the algorithm to be more greatly tested and determine potential limitations.

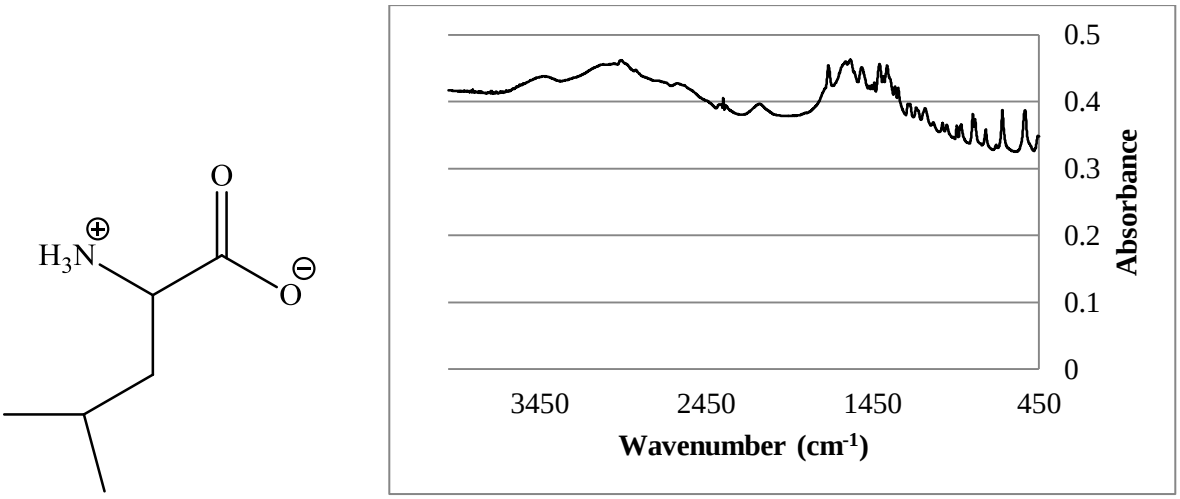


Figure 9. Structure and FT-IR spectrum of L-leucine.

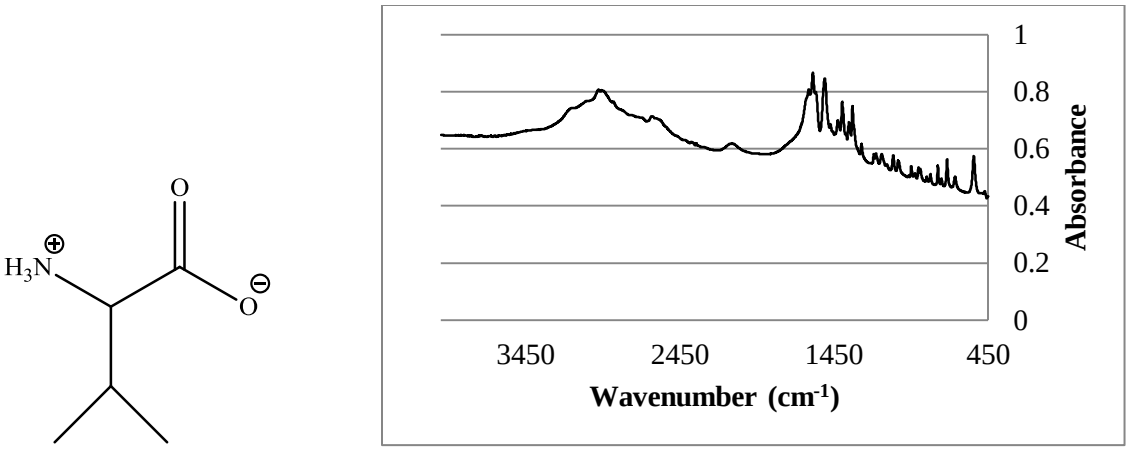


Figure 10. Structure and FT-IR spectrum of L-valine.

The initial standards were mixtures of the amino acids in various simple ratios. Two amino acids were used in a mixture set, but the composition of each amino acid would change between samples in the set. Below is an example of a tested dataset of glycine and L-leucine. Figure 11 shows the spectra of the mixtures and the percentage composition of each; for this set, the molar ratio changed by 25% with mixtures ranging from 100% glycine (Gly) to 100% L-leucine (Leu). The 50:50 mixture of glycine and L-leucine did saturate the instrument, but the mixture was still analyzed in order to determine initial success.

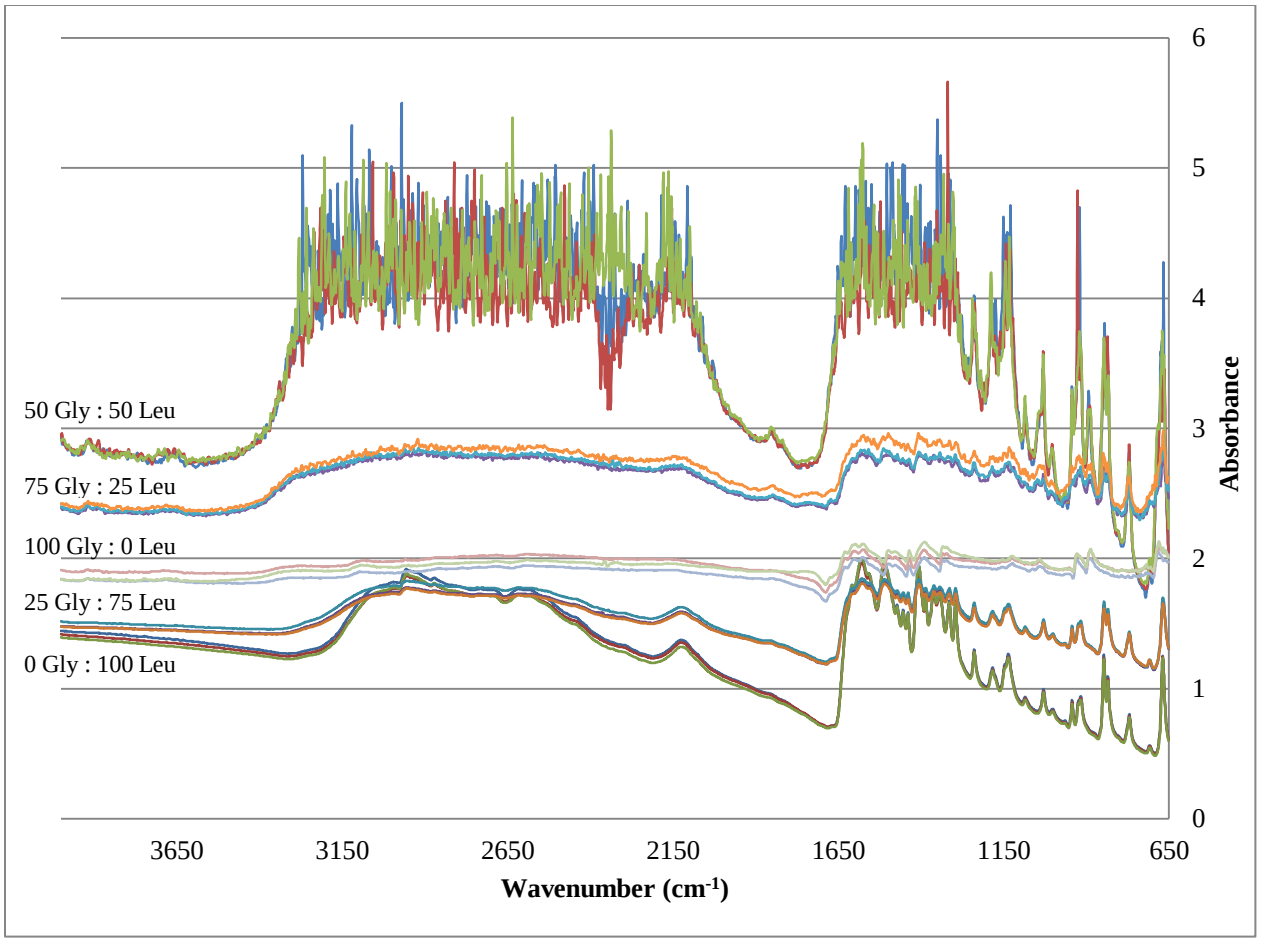


Figure 11. FT-IR spectra of glycine : L-leucine mixtures.

It should be noted that there are three spectra for each sample. The original procedure required pressing the amino acid mixture onto a sodium chloride salt plate with a microspatula; this procedure was used because the best solvent for many amino acids is water, which cannot be used as a solvent for IR spectroscopy. With this procedure, it was found that the layer of mixture on the salt plate would not always be even, despite best efforts. To account for the variability in the samples, three spectra of each sample were taken. Once a sample was "plated", a spectrum was taken; then the salt plate would be rotated approximately 120° and another spectrum was taken. The process was repeated to reach a full circle and yielding three spectra per sample.

When the spectra were analyzed with the algorithm, L-leucine was used as the POI by calculating the average of the three L-leucine spectra. Figure 12 shows the false-color image that is generated by the algorithm to portray the results, and the Mahalanobis distance values are presented in Figure 13. As it can be determined, Figures 12 and 13 consist of only the first PC, but Figure 12 shows that the program was successful in ranking the chemical composition in terms of similarity; the algorithm impressively ranked the 50 Gly : 50 Leu mixture correctly as well.

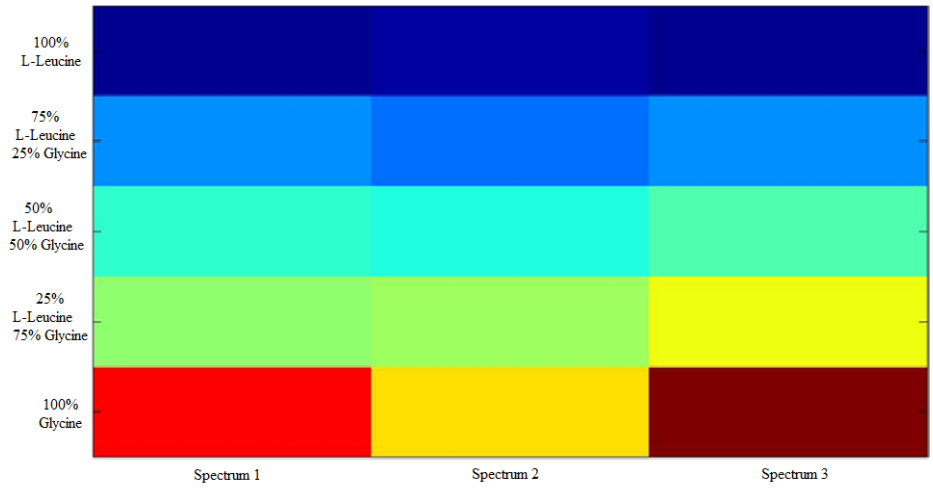


Figure 12. False-color image generated by the algorithm of glycine and L-leucine mixtures.

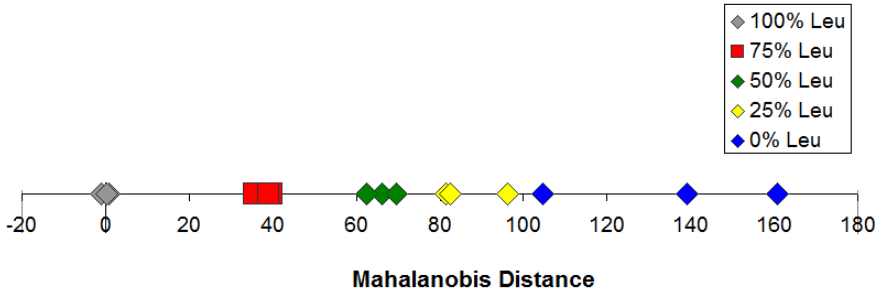


Figure 13. Plot of first PC of glycine and L-leucine mixtures.

The results given in Figures 12 and 13 were positive initial results. It was realized that the false color gradient did not best represent the chemical difference; it essentially showed the

chemical difference in one "dimension". To give an example, if collagen I, collagen III and elastin were inserted into the algorithm, and collagen I was the POI, the results would indicate that collagen III and elastin were both on the opposite end of the spectrum to show the difference from the POI. However, these results would also indicate that collagen III and elastin are similar, which in fact, they are not. Because of this problem, the algorithm was adjusted to place the POI in the center of the false color gradient, representing a light green color as shown in Figure 14. With this new configuration, collagen III and elastin would be on opposite ends of the spectrum, assuming that collagen I is still the POI. Following this initial test, it was determined that the data, or mixtures, needed to be more complex to further test the algorithm. The next step was to make more complicated mixtures and to synthesize an amino acid copolymer.

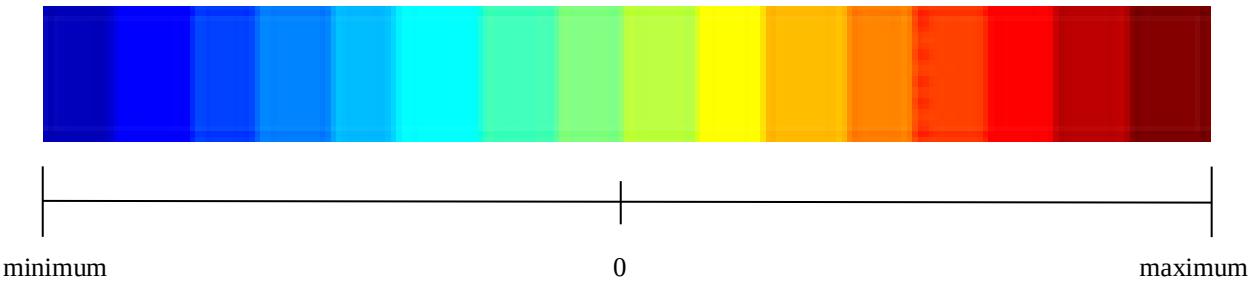


Figure 14. Modified false-color gradient of Mahalanobis distances.

The synthesis of an amino acid copolymer first enlisted the use of a microwave. A synthesis of polyglycine was previously achieved by allowing the glycine to react at 140 °C for 19 hours in aqueous ammonia.[13] The time length was not desirable, thus a similar procedure was created based on this synthesis; however, it employed the time-saving capabilities of a microwave reactor. The first reaction consisted of glycine and water held at 150 °C for ten minutes, but it was unsuccessful. This procedure was modified many times, as show in Table 1.[14] The various reaction conditions that were adopted from past syntheses were "converted" to a microwave reaction with the help of advisors and through the use of a research and development

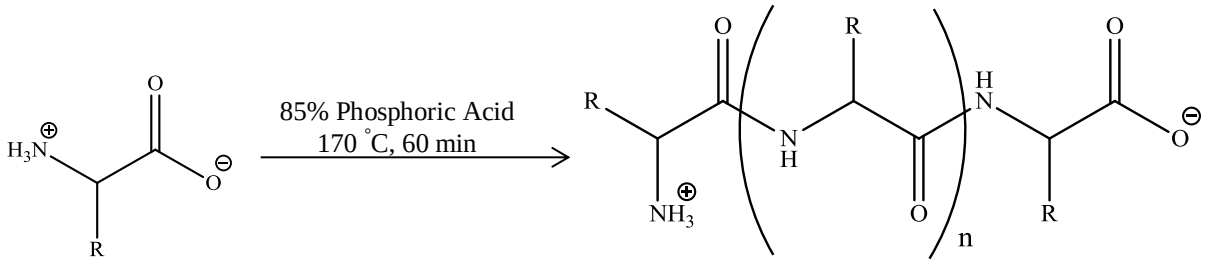
guide on continuous flow reactions and microwave reactions.[15] The name of each reaction correlates to the date of the reaction as well as the reaction conditions.

Table 1. Unsuccessful amino acid copolymer microwave reactions.

		Solvent Conditions		Microwave Conditions	
Name/Test	**Reagent Amount**	**Solvent**	**Catalyst**	**Time**	**Temp. (°C)**
2-6-12	0.5156 g Gly	10 mL deionized water (DI H_2O)	N/A	10 min	150
4-3-12 Plain	0.5626 g Gly	6.8 mL DI H_2O	N/A	10 min	150
4-3-12 w/HCL	0.5816 g Gly	6.6 mL DI H_2O	Drop of 12.1 N HCl	10 min	150
4-12-12 8mL	0.3991 g Gly	8 mL DI H_2O	N/A	10 min	150
4-12-12 6 mL	0.3635 g Gly	6 mL DI H_2O	N/A	10 min	150
4-12-12 4 mL	0.3670 g Gly	4 mL DI H_2O	N/A	10 min	150
4-12-12 2 mL	0.3577 g Gly	2 mL DI H_2O	N/A	10 min	150
5-31-12 2 mL SA	0.3500 g Gly	2 mL DI H_2O	Drop of 12 M H_2SO_4	1 hr	250
6-19-12 Test A	0.4074 g Gly	N/A	N/A	10 min	150
6-19-12 Test B	0.7566 g Gly	N/A	0.8 mL 85% H_3PO_4	10 min	160
6-20-12 Test C	0.4476 g Gly	2 mL THF	Drop of 9 N H_2SO_4	1 hr	250
6-27-12 Test B.2	0.7652 g Gly	N/A	0.8 mL 85% H_3PO_4	10 min	160
7-5-12 3G : 1L	0.4503 g Gly 0.2667 g Leu	N/A	0.8 mL 85% H_3PO_4	10 min	160
7-5-12 1G : 1L	0.3092 g Gly 0.5288 g Leu	N/A	0.8 mL 85% H_3PO_4	10 min	160
7-5-12 1G : 3L	0.1210 g Gly 0.6311 g Leu	N/A	0.8 mL 85% H_3PO_4	10 min	160
7-24-12 ~0.02 mol G	1.5019 g Gly	N/A	0.4 mL 85% H_3PO_4	15 min	170
7-25-12 Trial 8	1.0000 g Gly	0.1 mL DI H_2O	0.3 mL 85% H_3PO_4	15 min	170
7-25-12 Trial 11	1.0035 g Gly	N/A	0.3 mL 85% H_3PO_4	10 min	170
7-25-12 Trial 17	1.3328 g Gly	N/A	0.3 mL 85% H_3PO_4	15 min	160
7-25-12 Trial 18	1.1093 g Gly	N/A	0.3 mL 85% H_3PO_4	15 min	165
7-25-12 Trial 19	1.2409 g Gly	N/A	0.3 mL 85% H_3PO_4	10 min	165

Due to the lack of success with the microwave, a procedure with a reaction time of one hour was attempted following the exact procedure as stated in the literature. The reaction is given

in Scheme 1. Each "R" represents a specific side chain for each amino acid. For example, R for glycine is H.



Scheme 1. Amino acid thermal copolymerization in phosphoric acid.[16]

The first time this scheme was deemed successful, it was tested with three different mixtures and analyzed with both IR spectroscopy of and High-Performance Liquid Chromatography Time Of Flight Mass Spectrometry (HPLC-MS-TOF). The compositions of these mixtures are given in Table 2, including the mixtures that were repeated to ensure reproducibility after the initial successful reactions. The data for the "All Glycine" products are given in Figures 15, 16 and 17; the data for the remaining products can be found in Appendix B. It was at this point that a microwave reaction was eliminated and the reaction in Scheme 1 was used.

Table 2. Successful amino acid thermal copolymerization in phosphoric acid.

Name	**Reagent Amount**	**Amount of 85% H_3PO_4**
All Gly (July)	2.2525 g Gly	0.6 mL
2 Gly : 1 Leu (July)	1.5016 g Gly 1.3126 g Leu	0.6 mL
1 Gly : 2 Leu (July)	0.7512 g Gly 2.6249 g Leu	0.6 mL
All Gly (Aug)	1.1265 g Gly	0.3 mL
2 Gly : 1 Leu (Aug)	0.7506 g Gly 0.6589 g Leu	0.3 mL
1Gly : 2 Leu (Aug)	0.3772 g Gly 1.3107 g Leu	0.3 mL

The names of the mixtures correlate to the composition of the mixture as well as date of when the reaction was completed. While both reactions were completed roughly a week apart, the reactions occurred in different months, as indicated in the name. The IR spectra in Figure 15 show that the July and August products are identical and the product is indeed different from the glycine starting product. There are characteristic differences around 3300 cm^{-1} and the area below 1000 cm^{-1}. These differences are best explained by the broadening effect that occurs with increased hydrogen bonding, as expected with the copolymer product. Likewise, a dramatic increase in absorbance at 1650 cm^{-1} in these product spectra suggests the presence of a new amide carbonyl functional group, which is consistent with the expected copolymer product.

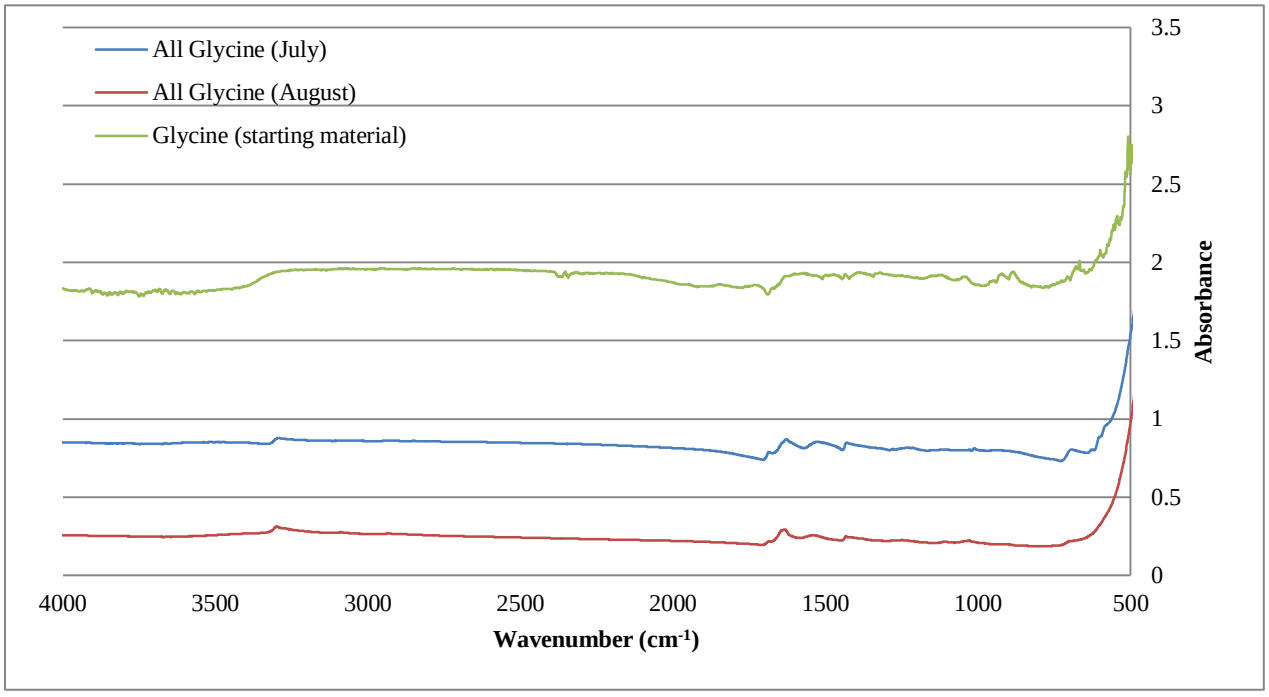


Figure 15. FT-IR spectra of "All Glycine" products and glycine starting material.

The data provided by the HPLC-MS-TOF also shows that the product is a copolymer of glycine. First, the retention time of glycine is 1.405 min. The retention times of this product are 1.545 minutes and 1.677 min, indicating that these products are larger and less polar than glycine alone. Additionally, the mass of the fragments correlate to very large chains. It should be noted

that the values 453 *m/z* and 359 *m/z*, are fragments that remained in the system from the 2 Gly : 1 Leu product, and the 215 *m/z* value is a separate compound that remained in the column. However, the value 286 *m/z* correlates to a five glycine unit chain. It is assumed that the smaller peaks are fragments of larger chains.

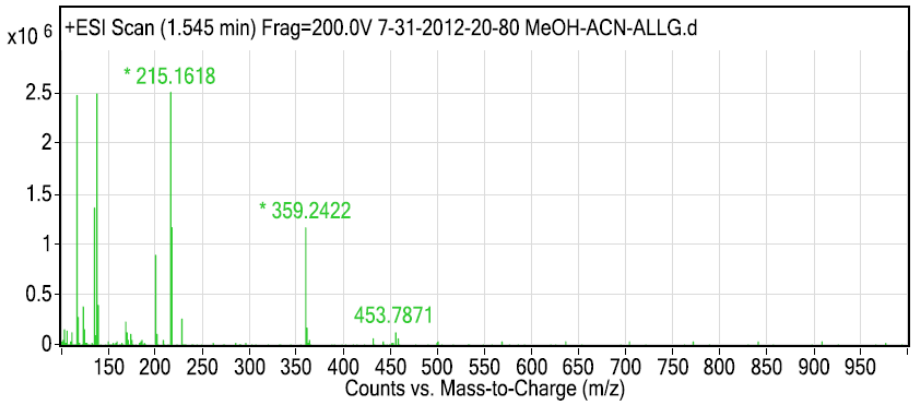


Figure 16. HPLC-MS-TOF All Glycine (July) data: retention time 1.545 minutes.

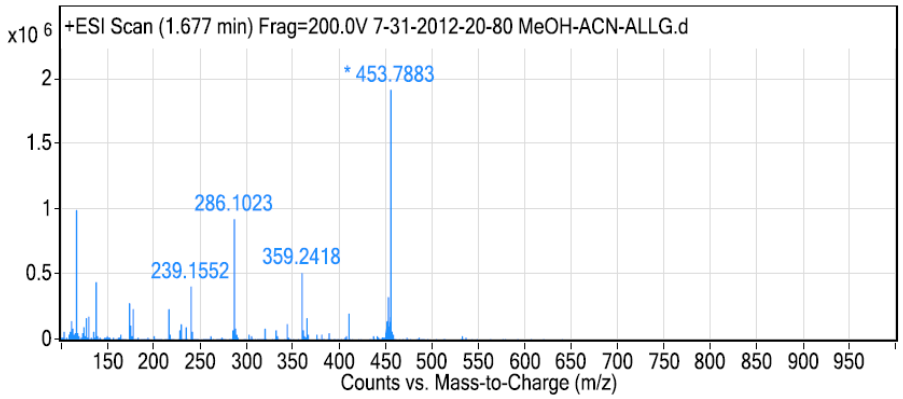


Figure 17. HPLC-MS-TOF All Glycine (July) data: retention time 1.677 minutes.

With a successful copolymer reaction established, the new standard mixtures could be created. As collagen is a known major component of the arterial wall, a polymer with a collagen-like distribution of amino acids would serve as an excellent model system in this study. The amino acid mixtures described were formulated and referred to as a "mock collagen" due to their ratio of amino acids present. Since the composition of collagen is one-third glycine and one-sixth L-proline, these molar ratios were maintained throughout all mixtures; the structure for each is

given below. The remaining fifty percent consisted of amino acids from the various subgroups: acidic, basic, polar, nonpolar, aromatic, and thiol. The list of amino acid mixtures is given in Table 3; the remaining amino acid structures are given in Appendix C.

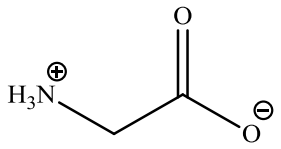


Figure 18. Glycine.

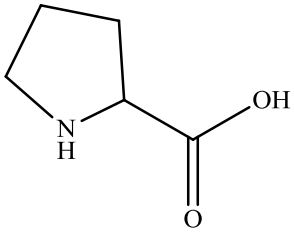


Figure 19. L-Proline.

Table 3. Mock collagen mixture composition.

Mixture Name	Amino Acid (s)	Composition
A	L-Valine L-Leucine	25% 25%
B	L-Valine L-Leucine	10% 40%
C	L-Valine L-Leucine	40% 10%
D	L-Tyrosine L-Valine	25% 25%
E	L-Glutamic Acid	50%
F	L-Glutamic Acid L-Valine	25% 25%
G	L-Glutamic Acid L-Leucine	25% 25%
H	L-Asparagine	50%
I	L-Asparagine L-Leucine	25% 25%
J	L-Asparagine L-Valine	25% 25%
K	L-Cysteine	50%
L	L-Cysteine L-Leucine	25% 25%
M	L-Cysteine L-Valine	25% 25%

The data collection procedure involved taking IR spectra of both the unreacted and reacted mixtures. First, each amino acid in a designated mixture was allocated and placed in a test tube. Once all of the amino acids were allocated, the mixture was generously ground together using a mortar and pestle made of agate. The mixture was then stored in an uncapped vial and placed in a vacuum desiccator to dry overnight. IR spectra of the unreacted mixtures were taken prior to the reaction using a FT-IR spectrometer. As previously described, the samples were pressed onto a sodium chloride salt plate by using a microspatula; however, this method presented several issues.

First, the sample would not spread easily on the plate; it would take a considerable amount of time to spread the sample which meant more time exposing it to air and water vapor. Next, the mixture would not spread evenly. In order to address this, three scans would be taken of a single sample, as previously explained. The faults of this original method were realized prior to creating the mock collagen mixtures, which allowed time to find a better method. It was determined that a potassium bromide press would be the best solution. Potassium bromide, like sodium chloride, does not absorb infrared light. The press uses a considerably smaller amount of sample for the IR scan, and it presses dry KBr with the sample so that the sample is actually contained within the salt plate. It effectively lowers the time and exposure to water vapor, yielding better, cleaner spectra. To ensure the quality of the spectra, the salt plate is still rotated 120° once to have two spectra of the same sample.

After the IR spectra had been collected of the unreacted mixtures, each mixture was reacted according to Scheme 1 to create "mock collagen" copolymers. Once cooled, the products were extracted from any starting material via water and dialysis tubing. The material was placed in dialysis tubing with a pore size diameter of 4.8 nm which holds molecular weights of 12,000 g and higher. The tubing was then placed in a beaker of deionized water with a magnetic stir bar to stir the water, with the water being refreshed every 20 minutes over a one hour period.

Very dry products were needed to create KBr pellets and obtain IR spectra of these products, but obtaining anhydrous solid products was non-trivial. Originally, the excess water would be evaporated by blowing a stream of gaseous nitrogen onto the product, and then it would be placed in the vacuum desiccator for two to three days. When reacted mixtures were analyzed with IR spectroscopy, the peak definition of each spectrum greatly diminished. The

products had a very gummy texture, meaning too much water was retained during the drying process, which made it difficult to achieve well-defined spectra.

In order to address this issue, a high vacuum drying apparatus was assembled, as shown in Figure 20. The sample, contained in a small, uncapped vial, is placed in the round-bottom flask (1). The vial is covered by a Kimwipe secured by wire in order to prevent the loss of product. The first round-bottom flask is connected to the second round-bottom flask via rubber tubing. The round-bottom flask (2) holds drying reagent phosphorus pentoxide, which is then connected a two-piece cold trap (3). The trap is submerged in a Dewar flask filled with dry ice and acetone. The final tube connects the cold trap to the high vacuum on the left. With minimum water residue in the sample upon commencing, it has been determined that it takes 6-8 hours within the drying apparatus to sufficiently dry a product.

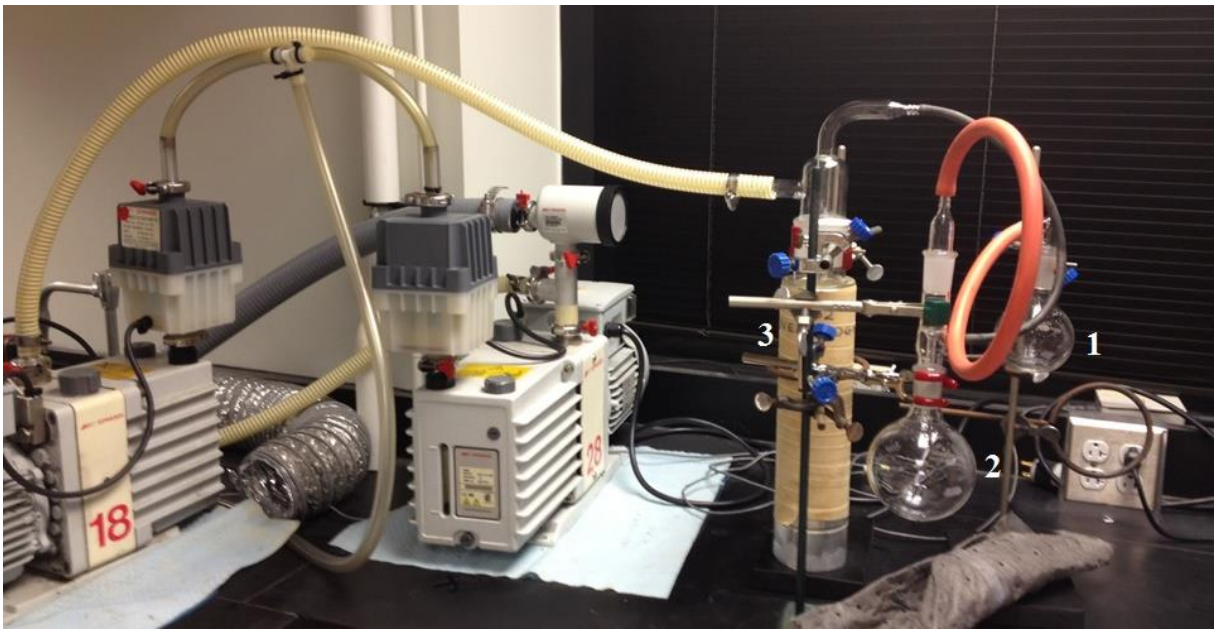


Figure 20. High-vacuum drying apparatus: 1 is a round-bottom flask containing sample; 2 is a round-bottom flask holding drying reagent; 3 is a two-piece cold trap submerged in a Dewar flask.

Validation of Chemical Algorithm:

The validation of the chemical algorithm was ideally to be completed with all of the mock collagen mixtures; however the delay in drying products only allowed for the L-asparagine mixtures to be analyzed. The mixtures compositions are given in Table 3, and they were made according to the method described above. Figure 21 shows the IR spectra of the unreacted mixtures, mixtures H, I and J. There are unexpected spectra features that appear between 2875 and 1875 cm^{-1} which is normally an unabsorbing region; it is likely that these features are a result of fringing and/or scattering. Because there are only six spectra (seven including the reference), PCA was not used for the analysis of these samples. It is suggested that PCA be used with a minimum sample size of fifty as according to the literature.[17] The reference spectra used for validation are a spectrum of the individual amino acid indicated. For each amino acid, two spectra were taken as according to the method previously described. These spectra were then averaged to be used as the reference.

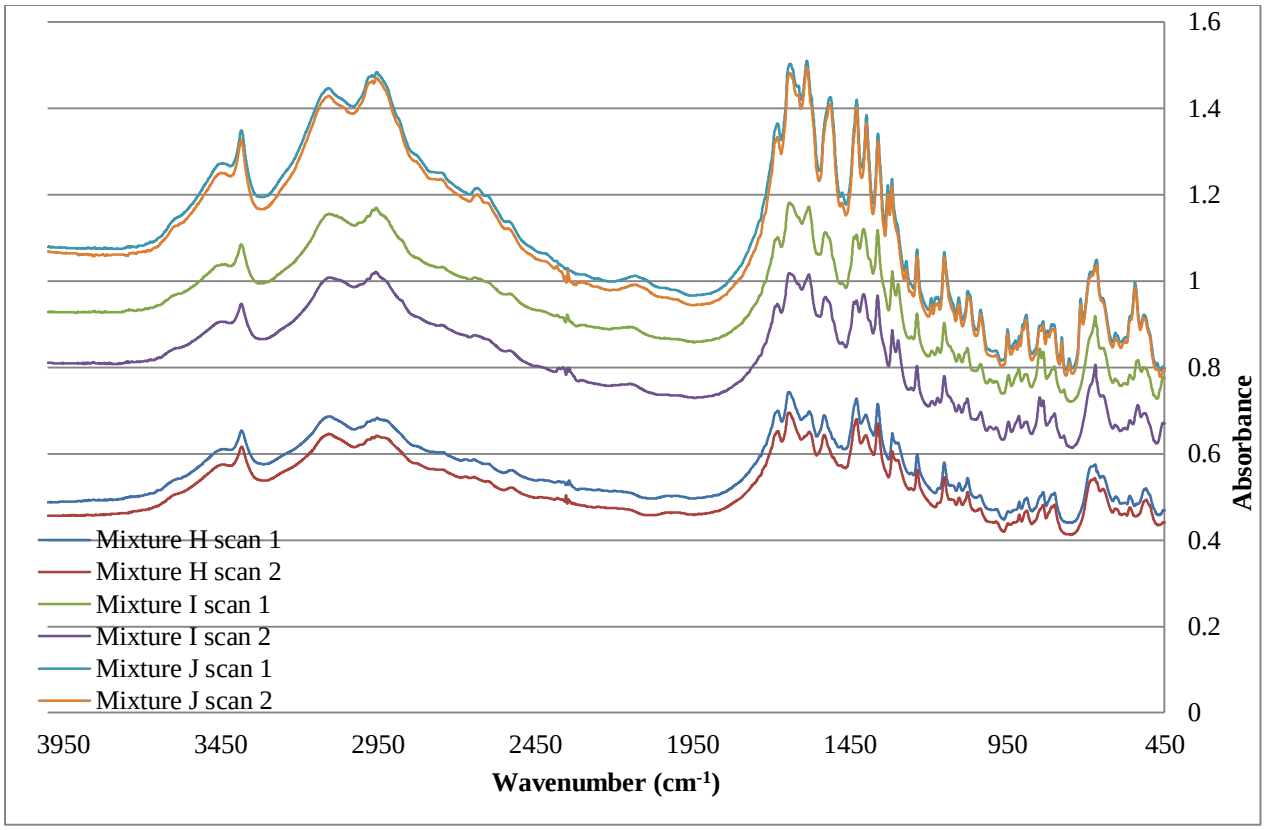


Figure 21. FT-IR spectra of unreacted mixtures H, I and J.

The unreacted mixtures H, I and J were then analyzed with the algorithm, and the results are given in Figure 22. The L-valine reference is the seventh column, because it is the light green color, it confirms that the reference was centered in the in false-color gradient. The mixture J spectra are considered "closest" to the reference based on the color scheme; this is also the expected result because mixture J contains L-valine and the other mixtures do not. It was predicted that mixture I would be the next "closest" to the L-valine reference because mixture I contains L-leucine. As mentioned before, L-valine and L-leucine have very similar IR spectra because there is only one carbon difference between the two molecules. These results confirm this prediction. Lastly, mixture H was considered to be the most chemical different because it contains neither L-leucine nor L-valine, and the results agree with this prediction. Glycine was also used as a reference for the different sets of spectra (Appendix E).

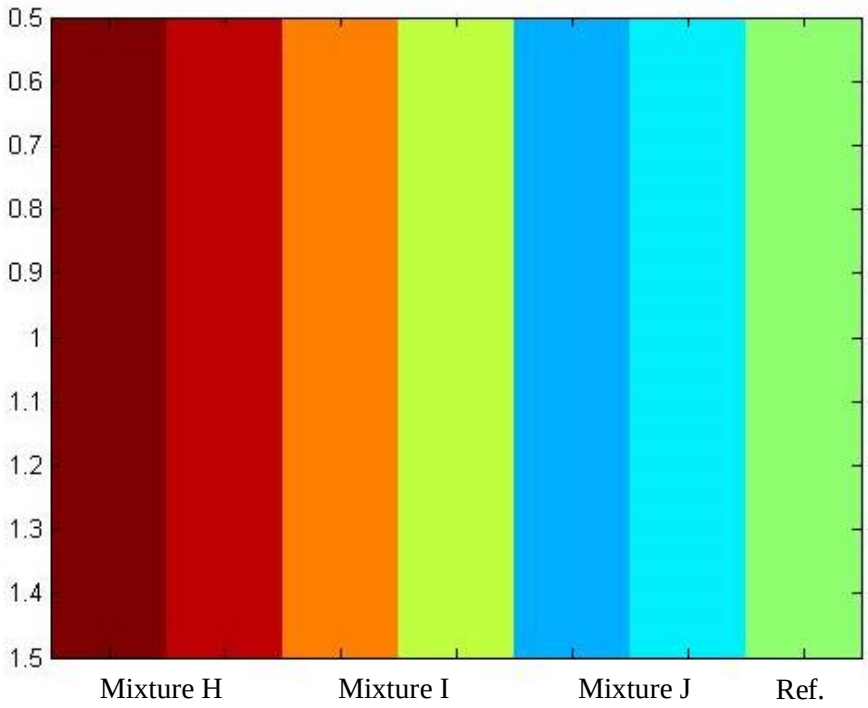


Figure 22. False-color image generated by the algorithm of unreacted mixtures H, I and J spectra: L-valine reference.

These preliminary results of the unreacted mixtures were very promising, but the reacted mixtures still needed to be analyzed. It can be seen in Figure 23 that the reacted mixtures have much less defined peaks in the spectra. This is due to the broadening effect that occurs with hydrogen bonds. To demonstrate the transformation between the unreacted and reacted mixtures, the differences between the mixture H spectra were calculated and shown in Figure 24. It shows that there is a considerable greater amount of character in the unreacted spectra, and there is an appreciable amount of hydrogen bonding to compress some of the large peaks, especially from 3100-2950 cm^{-1} and near 1650 cm^{-1}.

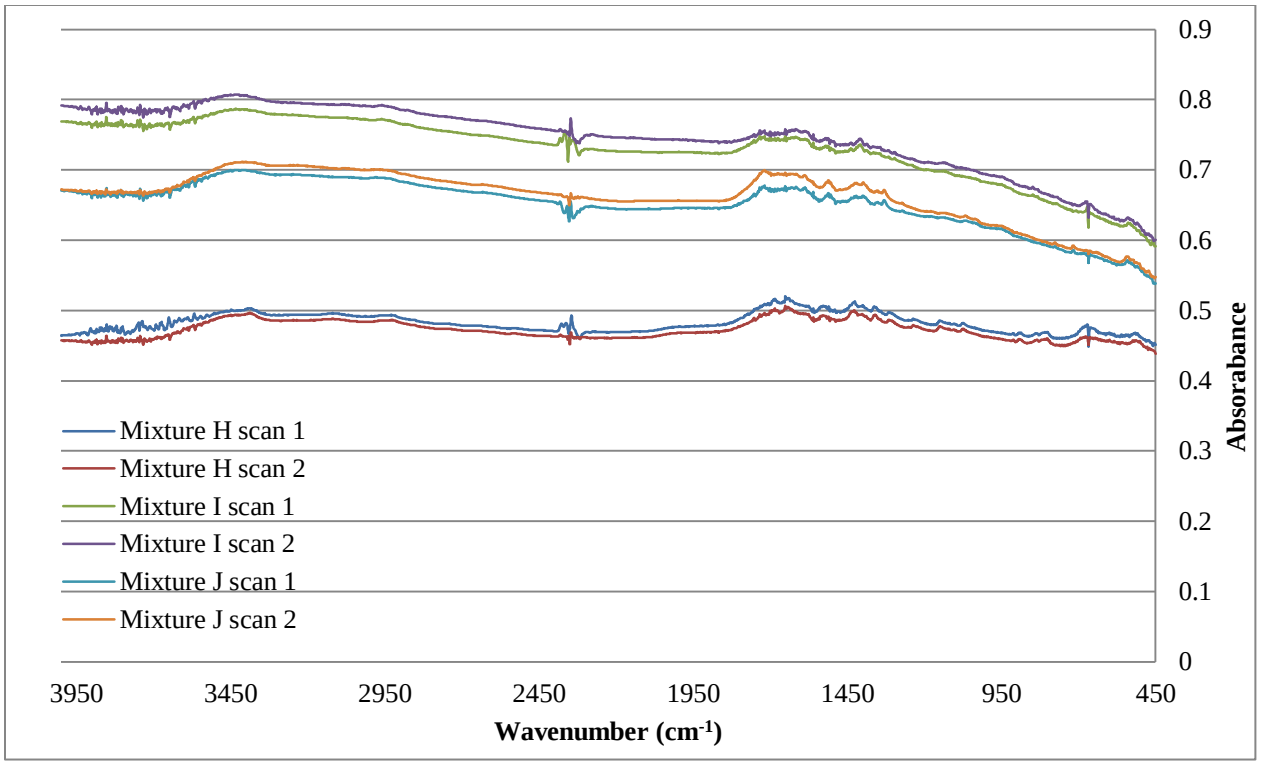


Figure 23. FT-IR spectra of reacted mixtures H, I and J.

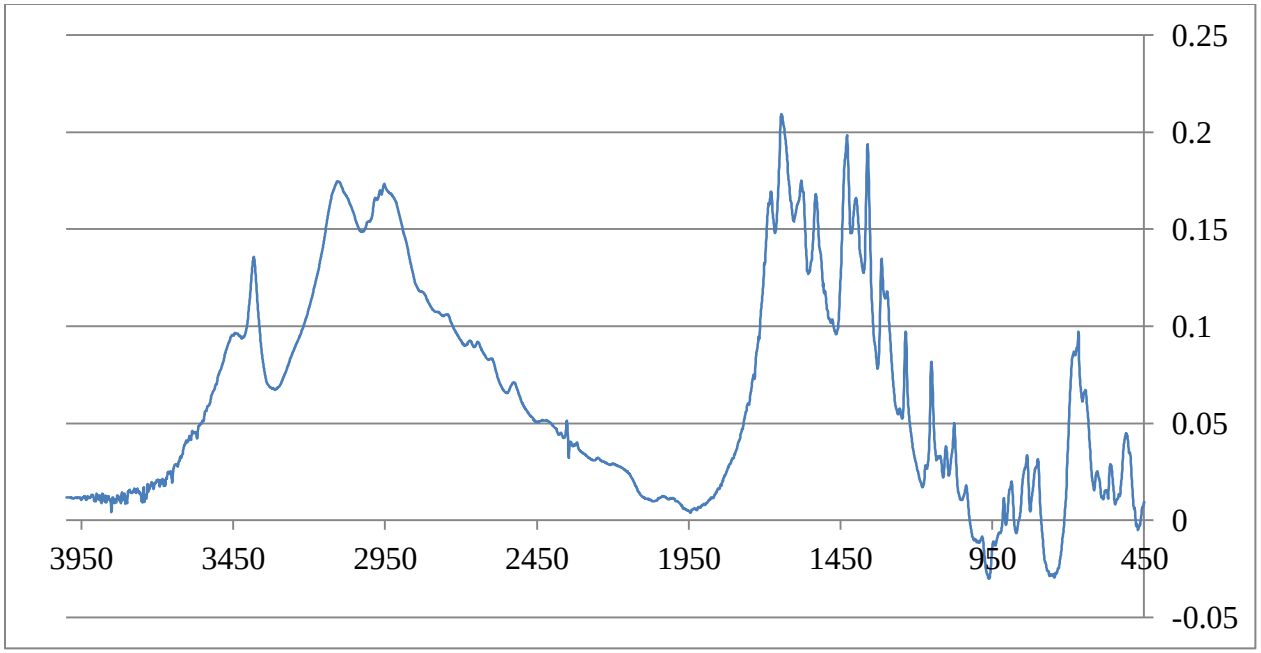


Figure 24. Difference of average unreacted mixture H and average reacted mixture H FT-IR spectra.

When the reacted mixtures spectra were analyzed with the algorithm, again using L-valine as a reference, the results (Figure 25) were similar to the results of the unreacted mixtures spectra. The mixture containing L-valine were most similar to the L-valine reference. Next in order was the mixture containing L-leucine, and the mixture that contained neither L-valine nor L-leucine was the furthest away, or least similar, to the L-valine reference. However, when glycine was used as the reference, the results were the same. It was determined that the spectra did not display enough variance for the algorithm to discern the true difference between the copolymers and the reference. Instead, the algorithm likely relied on the amount of absorbance of the reference (Figure 26).

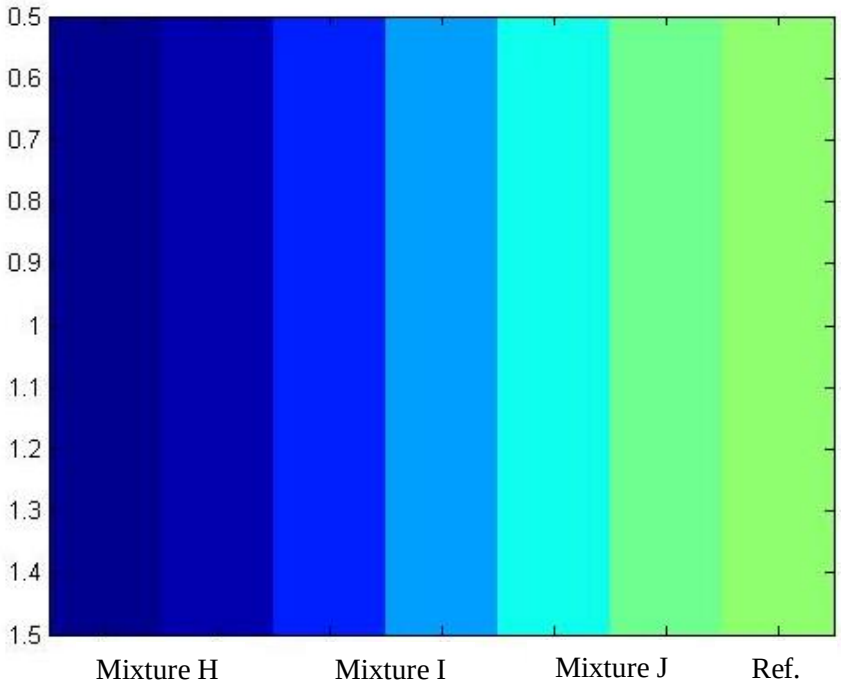


Figure 25. False-color image generated by the algorithm of reacted mixtures H, I and J spectra: L-valine reference.

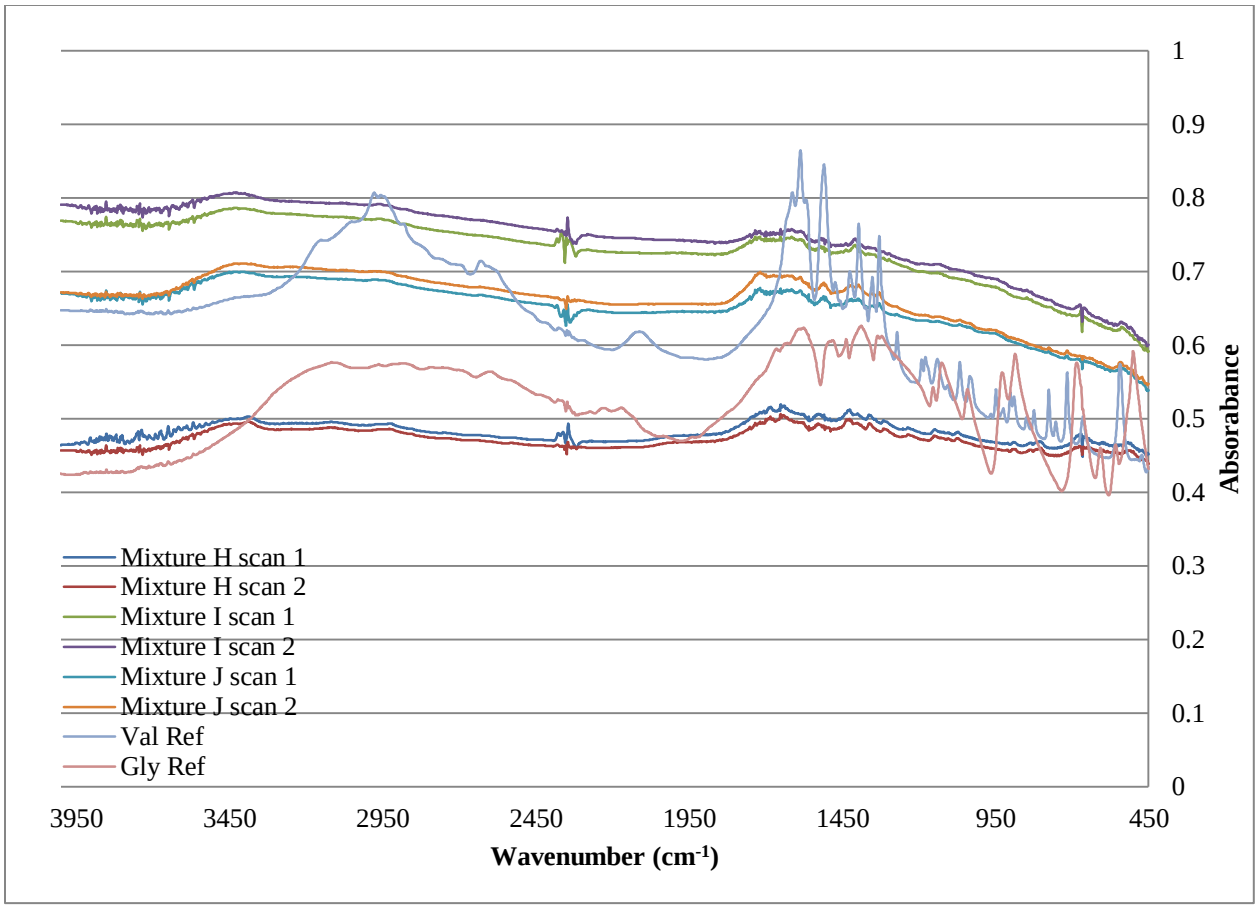


Figure 26. FT-IR spectra of reacted mixtures H, I and J with glycine and L-valine reference spectra.

In order to address this issue, the variance within the spectra needed to be highlighted. The greatest amount of variance existed between 1600 and 450 cm^{-1}, as represented in Figure 27. This selection was then used by the algorithm with L-valine as a reference. The results (Figure 28) show that the algorithm correctly ranked the mixtures, and these mixtures were specific to only L-valine. The results with glycine reference are provided in Appendix E.

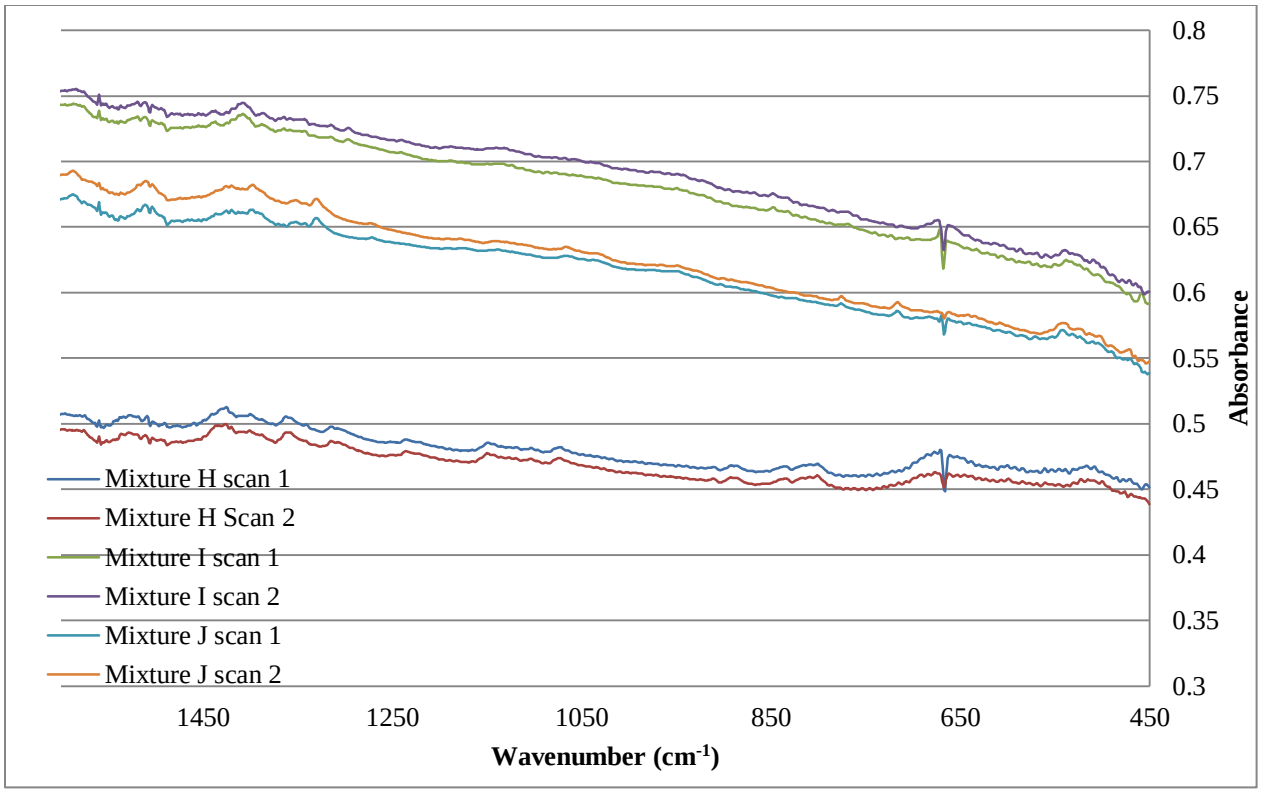


Figure 27. FT-IR selected spectra of reacted mixtures H, I and J.

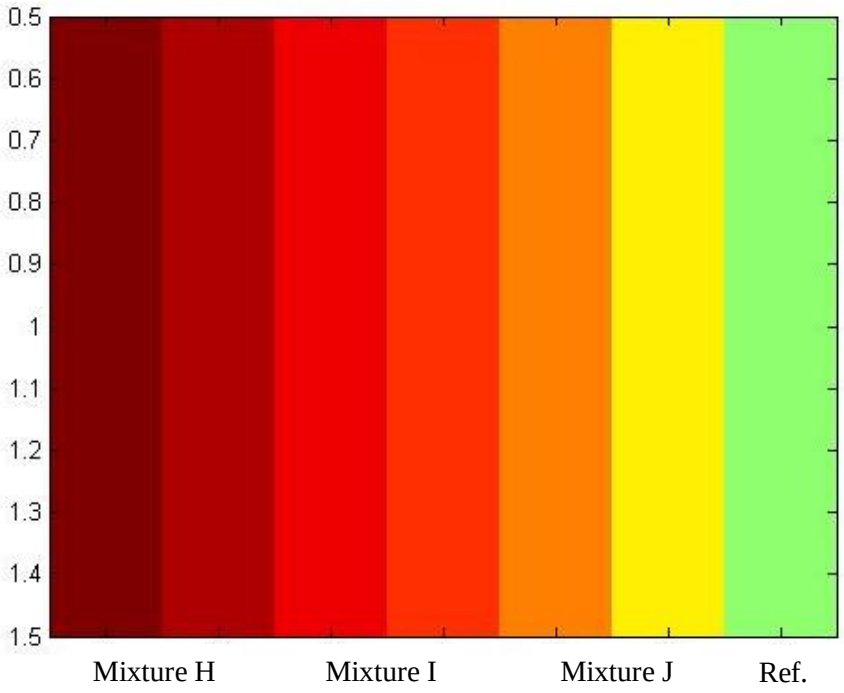


Figure 28. False-color image generated by the algorithm of reacted mixtures H, I and J selected spectra: L-valine reference.

Conclusion:

The initial results from the chemical algorithm support the hypothesis that the algorithm is working properly. It is able to accept a reference and use it as such for comparison. The algorithm results also suggest that PCA will be useful with larger dataset; as previously mentioned, it was inappropriate to use PCA due to the small number of spectra studied here. This rational is supported by the results shown in Figure 28. To obtain the results, the superfluous data was removed and a new dataset was created containing the largest variance of the original dataset. This is the purpose of PCA; therefore, when PCA is reinserted into the algorithm, it may prove to be functional.

At the moment, the algorithm is unfortunately not prepared for analysis of the aortic tissue due to lack of testable data. While it was intended to have validated the algorithm at this point, the delay in establishing a copolymer reaction as well as a drying mechanism has greatly slowed the progress. However, the initial results do prove promising based on the completed testing, specifically the differentiation between L-valine and L-leucine. Hopefully it will continue to prove functional as testing continues. When more data becomes available, the metrics of validation success will also be determined.

The algorithm was written specifically to analyze aortic tissue; however its function could be used also for other large datasets. Because the operations take little time to execute, it also could be used in a setting where large sets of data are collected rapidly, such as a hospital radiology department. Though it was written to determine a chemical difference between the healthy aorta and the aneurysm, it could potentially also be used in a diagnostic setting.

Validation of the algorithm will allow commencement of the analysis of the aortic tissue. Ideally the control data will be used to establish a norm for composition of the healthy aorta. As

many molecules have been previously identified to be contained within the aortic wall, the algorithm will assist in establishing a known concentration of these molecules. The “norm” will then be used against the aneurysm to determine both chemical differences but also concentration differences. Significant differences will then be considered for a chemical marker that is indicative of an AAA. Ideally the chemical marker would exist both in the aneurysm and blood, and therefore it could be detected through a blood test. Once this chemical marker is established, the overall objective of this project may be achieved.

Appendix A. MATLAB® Functions

Contents

Function A1 – prepdata

```
function data = prepdata (data)

if min(min(data)) >= 0 && max(max(data)) >= 20,
    if min(min(data)) == 0,
        data = (data + 0.0001);
    end
    data = log(100./data);
end

if min(min(data)) <= 0,
    data = (data - min(min(data)) + 0.0001);
end

data = mscorrect(data);
data = zscore(data);

assignin('base','zabsdata',data)



end
```

Function A2 – optPCA

```
function scores=optPCA(data,lv)

%PCA Principal components analysis optimized
%  PCA uses svd to perform pca on a data matrix. It is
%  assumed that samples are rows and variables are columns.
%  The input is the data matrix (data). Outputs are the scores
%  (scores) and loadings (loads).
tic
a = 1;
[m,n] = size(data);

  if (lv > min([m n])) %& (plots ~= 0)
    disp('Resetting lv to be equal to min([m n])')
    lv = min([m n]);
  end
toc
if n < m
    cov = (data'*data)/(m-1);
    [u,s,v] = svd(cov);
    assignin('base','u',u)  %   outputs u to the workspace
    assignin('base','s',s)  %   outputs s to the workspace
    assignin('base','v',v)  %   outputs v to the workspace
    u = 0;
    s = 0;
    cov = 0;

    toc


    %  Form the PCA Model Based on the Number of PCs Chosen
    loads  = v(:,1:lv);
    toc
    scores = data*loads;
    toc
else,
    [scores] = pca(data,0,0,lv);
end

assignin('base','scores',scores)    %   outputs scores to the workspace



%figure
%plot3 (scores(:,1),scores(:,2),scores(:,3),'b.');

%xlabel('PC1');
%ylabel('PC2');
%zlabel('PC3');
```

Function A3 – massdat

```
function [ output_args ] = massdat(rawdata,amountPCs,poi,width,height)
%massdat is able to analyze a large quantity of data. It uses PCA and
%Mahalanobis distances to decrease the amount of data based on variance,
%then compare it to a reference, respectively.
%   rawdata is the dataset to be analyzed
%amountPCs is the amount of PCs that are to be included in the calculation
%poi is the reference. The reference should be inserted as the final
%column in the dataset

correcteddata=prepdata(rawdata);%correct raw data
p=(width)-1;
scores=optPCA(correcteddata,p);%principal component analysis of corrected raw
data; p principal components calculated
arraylength=length(rawdata);%total number of pixels in image
arraylength=width;
if amountPCs==1%only 1 principal component desired
    R=input('Which PC?  ');
   pc=scores(:,R);%extract a column of scores from array generated by optPCA
   %pc=correcteddata(:,R); %This line is available for small dataset (N<50)
and should be inserted
   %when PCA is not being used.
   stdn=std(pc);%standard deviation of data points in specified PC
    for i=1:1:arraylength%total number of points
        mahaldist=(1/stdn)*(pc(i,:)-pc(poi,:));%mahalanobis
distance=euclidean distances/standard deviation
        if i==1
            mahalanobisdistances=mahaldist;%create array
        else
            mahalanobisdistances=[mahalanobisdistances;mahaldist];%add to
array
        end
    end
    assignin('base','mahalanobisdistances',mahalanobisdistances);
   sortedmahalanobisdistances=sort(mahalanobisdistances);%sort
   assignin('base','sortedmahalanobisdistances',sortedmahalanobisdistances);
   minimum=min(sortedmahalanobisdistances);%This centers the false-color
scheme center iteself around the poi
   maximum=max(sortedmahalanobisdistances);
   if abs(minimum)>maximum%if the absolute value of the minimum distance
value is greater than the maximum
         clim(1,1)=minimum;%minimum value becomes lower limit
         clim(1,2)=-minimum%opposite minimum value becomes upper limit
   else clim(1,1)=-maximum;%otherwise, opposite max becomes lower limit
       clim(1,2)=maximum;%max becomes upper limit
   end
   imagedata=reshape(mahalanobisdistances,width,1);
   imagedata=imagedata';
   assignin('base','imagedata',imagedata);
   imagesc(imagedata,clim);
      colormap(jet)
else if amountPCs>=2
        %option1%indicate pcs by user interface control object
        %option2
        %firstPC=scores(:,option1);%column in scores
```

```
        %secondPC=scores(:,option2);%column in scores
        %std1=STD(firstPC(:,1));%column standard deviation
        %std2=STD(secondPC(:,1));%column standard deviation
      %  R=input('Which PCs');
      %  if R=='all'
      %    n=1:1:amountPCs;
      %    assignin('base','n',n);
      %  else if R~='all'
      %          n=R;
      %      end
      %  end
      %  for n=1:1:amountPCs%total number of PCs

      reply=input('All PCs? Y/N [Y]:','s');
         if isempty(reply)
             reply='Y';
            n=1:1:p;
            %n=n';
           else reply='N'; %Only the PCs selectively entered will be
compared.
              chosenPCs=input('Enter a variable');
              chosenPCs=chosenPCs';
              assignin('base','chosenPCs',chosenPCs)
              %chosenPCs=keyboard('Which PCs?')
              y=length(chosenPCs);
              n=(scores(:,chosenPCs));
             % n=n';

         end

         for i=1:1:arraylength%total number of points
            PCn=scores(:,n);
            assignin('base','PCn',PCn)
            stdn=std(PCn(:,n));
            mahaldistn=(1/stdn)*(PCn(i,1)-PCn(poi,1));%column mahalanobis
distances
             if i==1
            mahalanobisdistances=mahaldistn;%create array
             else
                mahalanobisdistances=[mahalanobisdistances;mahaldistn];%add
to array
             end
             assignin('base','mahalanobisdistances',mahalanobisdistances)
         end

             %for j=1:1:arraylength%total number of points
        %    mahaldist2=(1/std2)*(secondPC(j,1)-secondPC(poi,1));%column
mahalanobis distances
        %     if j==1
        %    mahalanobisdistances2=mahaldist2;%create array
        %     else
        %
mahalanobisdistances2=[mahalanobisdistances2;mahaldist2];%add to array
        %    end
        %end
                for x=1:1:arraylength%total number of points
```

```
            vectorsum=sqrt(sum((mahalanobisdistances(x,1))^2));%sum
components
                if x==1
            vectorsums=vectorsum;%create array
                 else
                vectorsums=[vectorsums;vectorsum];%add to array
                end
                end

   assignin('base','vectorsums',vectorsums);
   sortedvectorsums=sort(vectorsums);%sort
   %clim(1,1)=sortedvectorsums(round(0.02*arraylength));%lower limit
   %clim(1,2)=sortedvectorsums(round(0.98*arraylength));%upper limit
   minimum=min(sortedvectorsums);%This centers the false-color scheme center
iteself around the poi
   maximum=max(sortedvectorsums);
   if abs(minimum)>maximum%if the absolute value of the minimum distance
value is greater than the maximum
          clim(1,1)=minimum;%minimum value becomes lower limit
          clim(1,2)=-minimum%opposite minimum value becomes upper limit
   else clim(1,1)=-maximum;%otherwise, opposite max becomes lower limit
       clim(1,2)=maximum;%max becomes upper limit
   end
   imagedata=reshape(vectorsums,width,height);
   imagedata=imagedata';
   imagesc(imagedata,clim);
   colormap(jet)



   % else%if amountPCs==3
   %     option1
   %     option2
   %     firstPC=scores(:,option1);%column in scores
   %     secondPC=scores(:,option2);%column in scores
   %     thirdPC=scores(:,option3);%column in scores
   %     std1=STD(firstPC(:,1));%column standard deviation
   %     std2=STD(secondPC(:,1));%column standard deviation
   %     std3=STD(thirdPC(:,1));%column standard deviation
   %     for i=1:1:arraylength%total number of points
    %        mahaldist1=(1/std1)*(firstPC(i,1)-firstPC(poi,1));%column
mahalanobis distances
    %         if i==1
    %        mahalanobisdistances1=mahaldist1;%create array
    %         else
    %
mahalanobisdistances1=[mahalanobisdistances1;mahaldist1];%add to array
    %        end
    %    end
    %    for j=1:1:arraylength%total number of points
    %        mahaldist2=(1/std2)*(secondPC(j,1)-secondPC(poi,1));%column
mahalanobis distances
    %         if j==1
    %        mahalanobisdistances2=mahaldist2;%create array
    %         else
```

```
    %
mahalanobisdistances2=[mahalanobisdistances2;mahaldist2];%add to array
    %         end
    %    end
    %    for k=1:1:arraylength
    %        mahaldist3=(1/std3)*(thirdPC(k,1)-thirdPC(poi,1));%column
mahalanobis distances
    %        if k==1
    %            mahalanobisdistances3=mahaldist3;%create array
    %        else
    %
mahalanobisdistances3=[mahalanobisdistances3;mahaldist3];%add to array
    %        end
    %    end
    %    for x=1:1:arraylength%total number of points
    %
vectorsum=sqrt(((mahalanobisdistances1(x,1))^2)+((mahalanobisdistances2(x,1))
^2)+((mahalanobisdistances3(x,1))^2));%sum components
    %        if x==1
    %        vectorsums=vectorsum;%create array
    %        else
    %            vectorsums=[vectorsums;vectorsum];%add to array
    %        end
    %    end
 %  assignin('base','vectorsums',vectorsums);
 %  sortedvectorsums=sort(vectorsums);%sort
 %  clim(1,1)=sortedvectorsums(round(0.02*arraylength));%lower limit
 %  clim(1,2)=sortedvectorsums(round(0.98*arraylength));%upper limit
 %  imagedata=reshape(vectorsums,height,width);
 %  imagesc(imagedata,clim);
 %  colormap(jet)
 %  end

     end
end

end
```

Function A4 – reshape

```
%RESHAPE Reshape array.
%   RESHAPE(X,M,N) returns the M-by-N matrix whose elements
%   are taken columnwise from X.  An error results if X does
%   not have M*N elements.
%
%   RESHAPE(X,M,N,P,...) returns an N-D array with the same
%   elements as X but reshaped to have the size M-by-N-by-P-by-...
%   M*N*P*... must be the same as PROD(SIZE(X)).
%
%   RESHAPE(X,[M N P ...]) is the same thing.
%
%   RESHAPE(X,...,[],...) calculates the length of the dimension
%   represented by [], such that the product of the dimensions
%   equals PROD(SIZE(X)). PROD(SIZE(X)) must be evenly divisible
%   by the product of the known dimensions. You can use only one
%   occurrence of [].
%
%   In general, RESHAPE(X,SIZ) returns an N-D array with the same
%   elements as X but reshaped to the size SIZ.  PROD(SIZ) must be
%   the same as PROD(SIZE(X)).
%
%   See also SQUEEZE, SHIFTDIM, COLON.

%   Copyright 1984-2005 The MathWorks, Inc.
%   $Revision: 5.14.4.5 $  $Date: 2009/05/18 20:48:03 $
%   Built-in function.
```

Function A5 – imagesc

```
function h = imagesc(varargin)
%IMAGESC Scale data and display as image.
%   IMAGESC(...) is the same as IMAGE(...) except the data is scaled
%   to use the full colormap.
%
%   IMAGESC(...,CLIM) where CLIM = [CLOW CHIGH] can specify the
%   scaling.
%
%   See also IMAGE, COLORBAR, IMREAD, IMWRITE.
 %   Copyright 1984-2005 The MathWorks, Inc.
%   $Revision: 5.11.4.5 $

clim = [];
switch (nargin),
  case 0,
    hh = image('CDataMapping','scaled');
  case 1,
    hh = image(varargin{1},'CDataMapping','scaled');
  case 3,
    hh = image(varargin{:},'CDataMapping','scaled');
  otherwise,

    % Determine if last input is clim
    if isequal(size(varargin{end}),[1 2])
      str = false(length(varargin),1);
      for n=1:length(varargin)
        str(n) = ischar(varargin{n});
      end
      str = find(str);
      if isempty(str) || (rem(length(varargin)-min(str),2)==0),
        clim = varargin{end};
        varargin(end) = []; % Remove last cell
      else
        clim = [];
      end
    else
      clim = [];
    end
    hh = image(varargin{:},'CDataMapping','scaled');
end

% Get the parent Axes of the image
cax = ancestor(hh,'axes');

if ~isempty(clim),
  set(cax,'CLim',clim)
elseif ~ishold(cax),
  set(cax,'CLimMode','auto')
end

if nargout > 0
    h = hh;
end
```

Appendix B. Glycine and L-Leucine Copolymer FT-IR Spectra and HPLC-MS-TOF Data

Glycine Retention Time: 1.405 min

L-Leucine Retention Time: 1.497 min

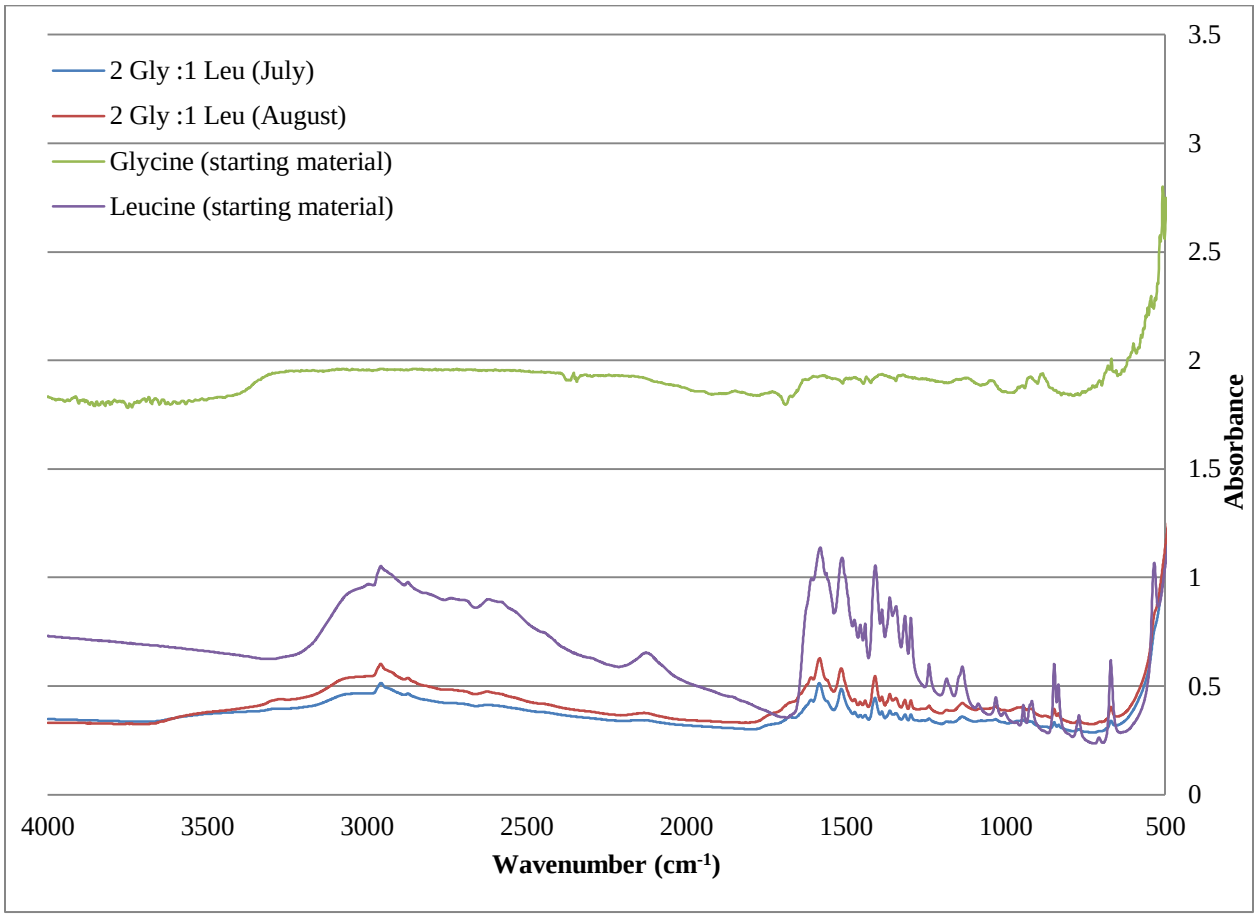


Figure B1. FT-IR spectra of 2 Gly : 1 Leu copolymers and starting reagents.

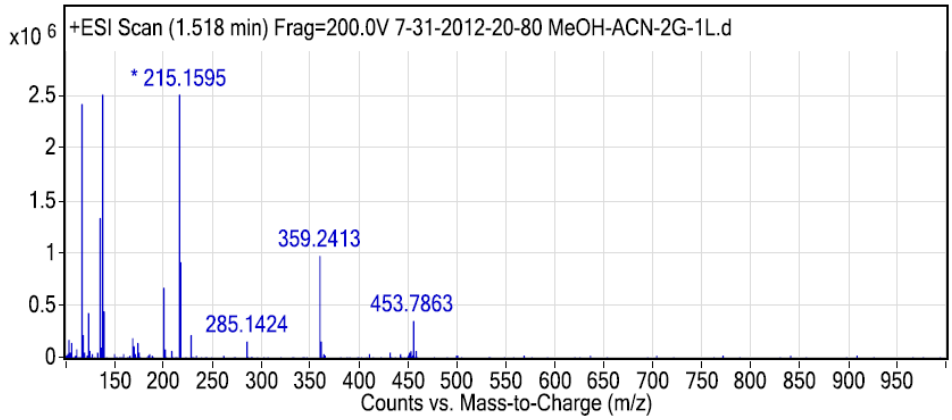


Figure B2. HPLC-MS-TOF 2 Gly : 1 Leu (July) data: retention time 1.518 minutes. 359 *m/z* correlates to a 3 Gly and 11 Leu chain.

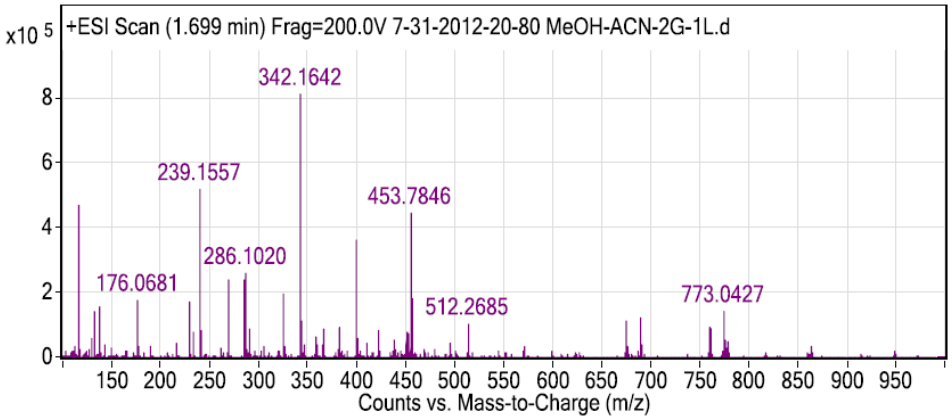


Figure B3. HPLC-MS-TOF 2 Gly : 1 Leu (July) data: retention time 1.699 minutes. 176 *m/z* correlates to an 8 Gly and 2 Leu chain.

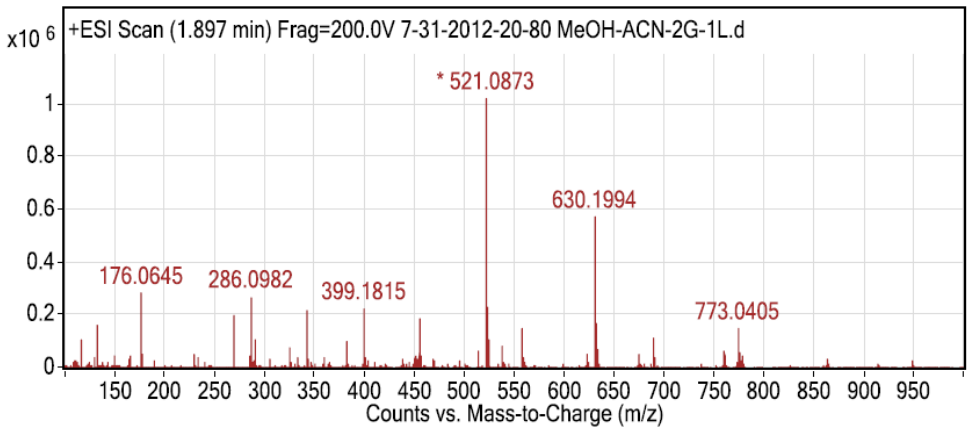


Figure B4. HPLC-MS-TOF 2 Gly: 1 Leu (July) data: retention time 1.897 minutes. 176 *m/z* correlates to an 8 Gly and 2 Leu chain.

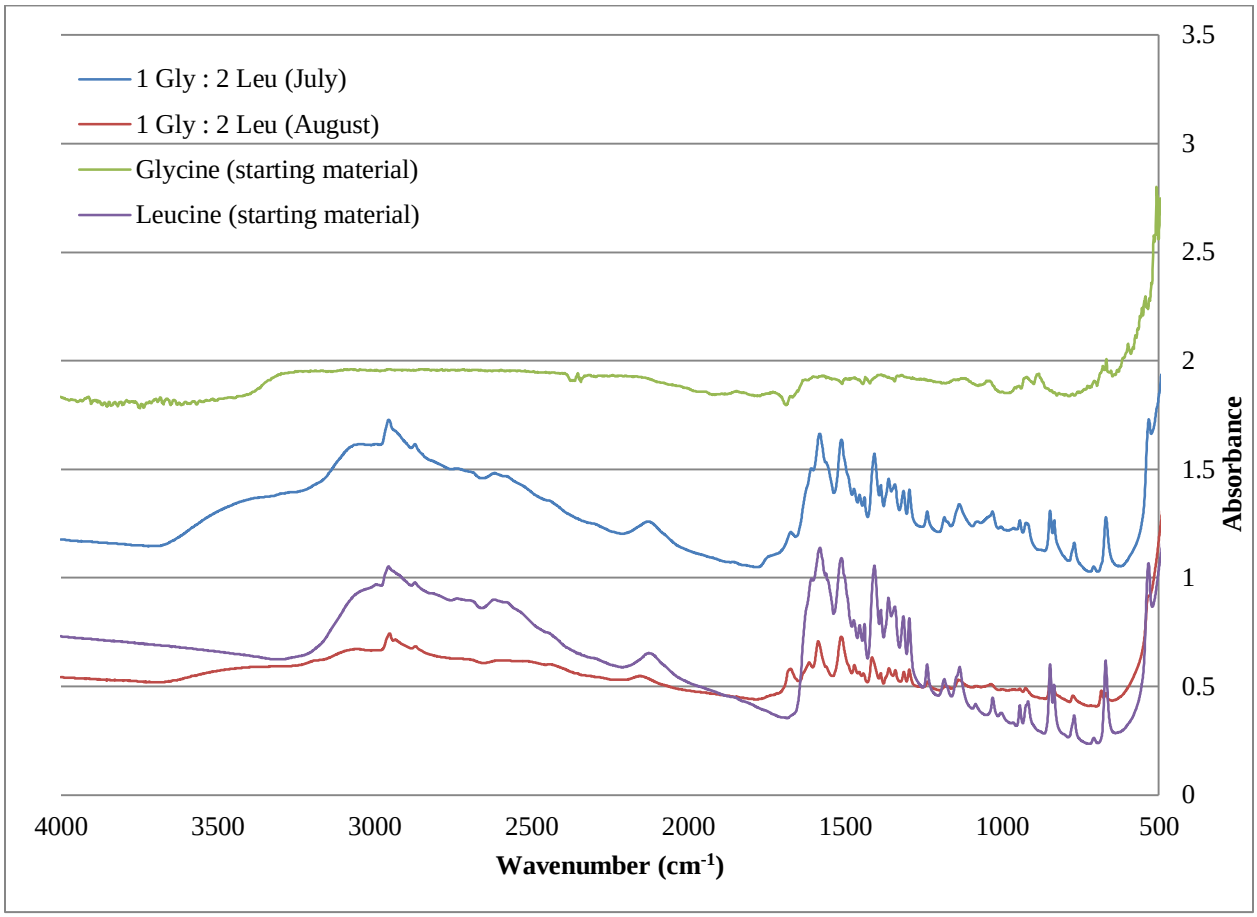


Figure B5. FT-IR spectra of 1 Gly : 2 Leu copolymers and starting reagents.

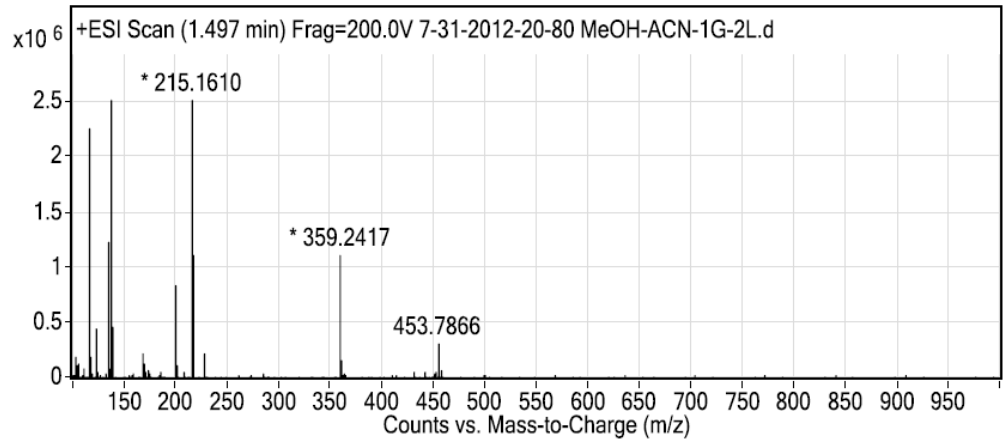


Figure B6. HPLC-MS-TOF 1 Gly: 2 Leu (July) data: retention time 1.497 minutes. 105 *m/z* correlates to a 6 Gly and 1 Leu chain.

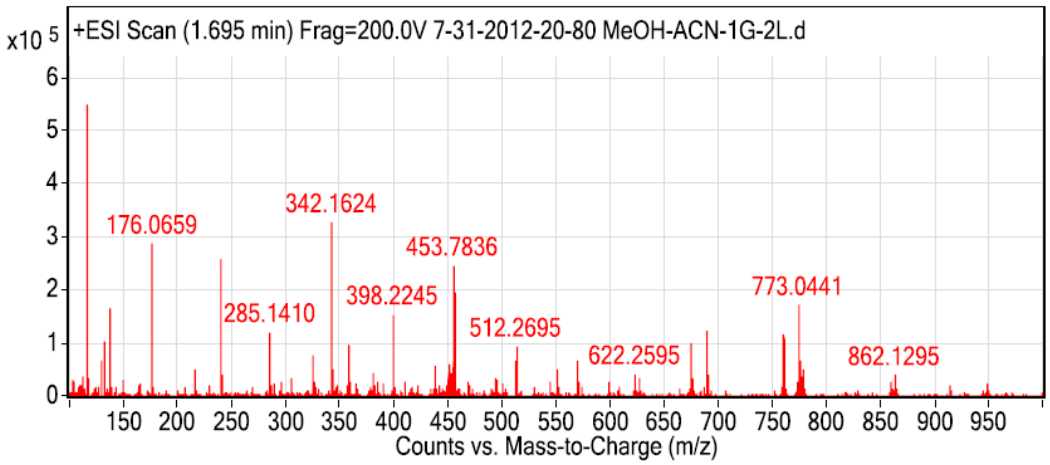


Figure B7. HPLC-MS-TOF 1 Gly: 2 Leu (July) data: retention time 1.695 minutes. 176 *m/z* correlates to an 8 Gly and 2 Leu chain.

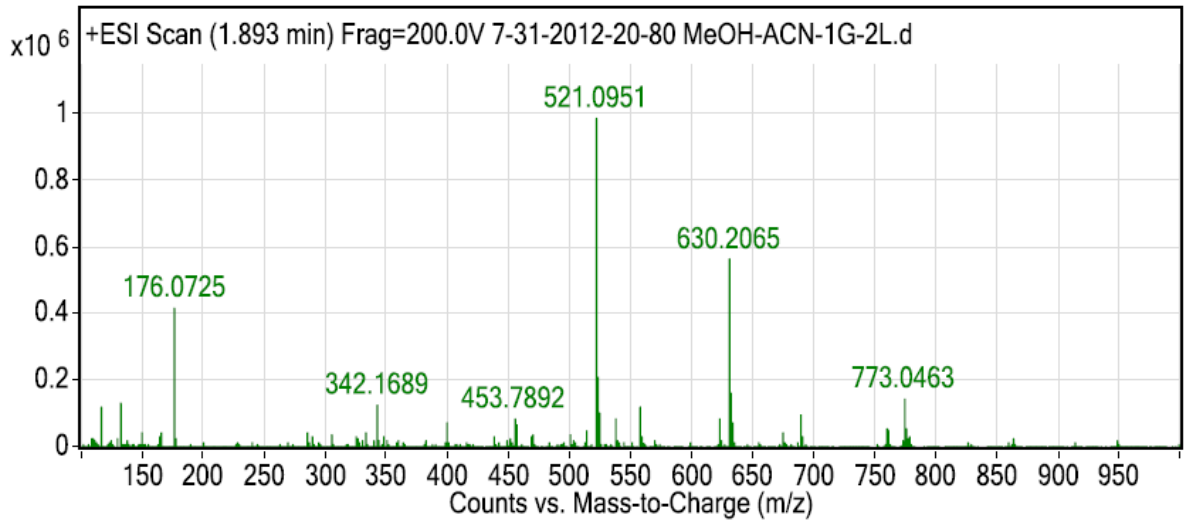
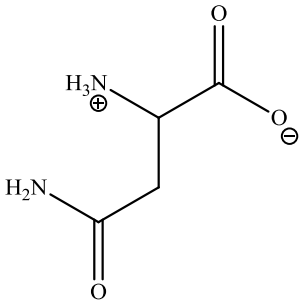


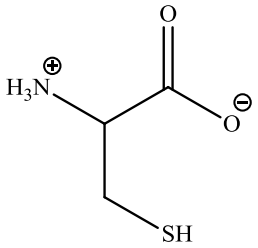
Figure B8. HPLC-MS-TOF 1 Gly: 2 Leu (July) data: retention time 1.893 minutes. 774 *m/z* correlates to a 3 Gly and 12 Leu chain.

Appendix C. Amino Acids

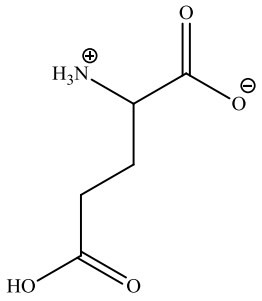
L-Asparagine



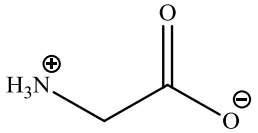
L-Cysteine



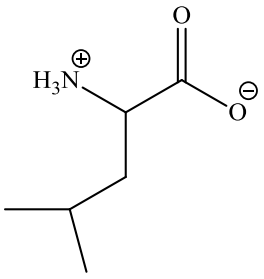
L-Glutamic Acid



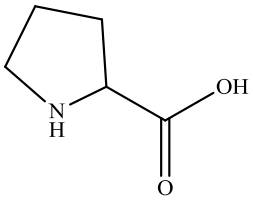
Glycine



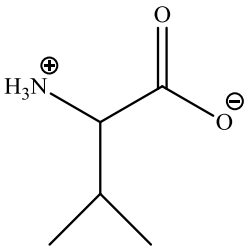
L-Leucine



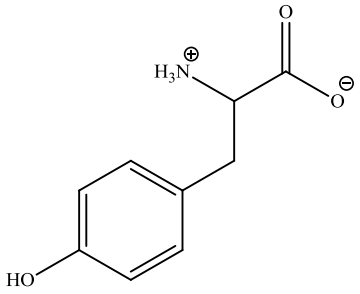
L-Proline



L-Valine



L-Tyrosine



Appendix D. Unreacted Mixture FT-IR Spectra

For mixture compositions, see Table 3.

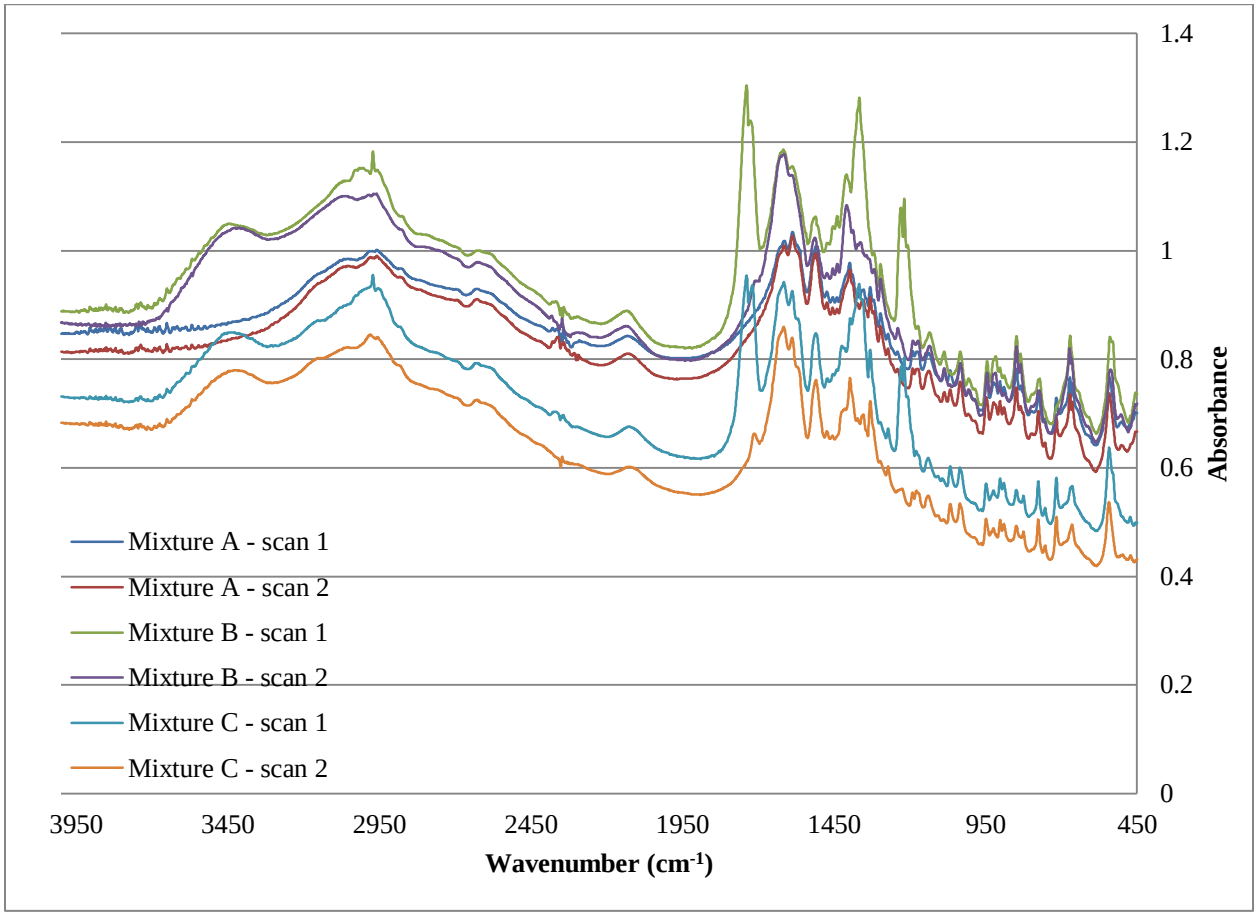


Figure D1. FT-IR spectra of unreacted mixtures A, B and C.

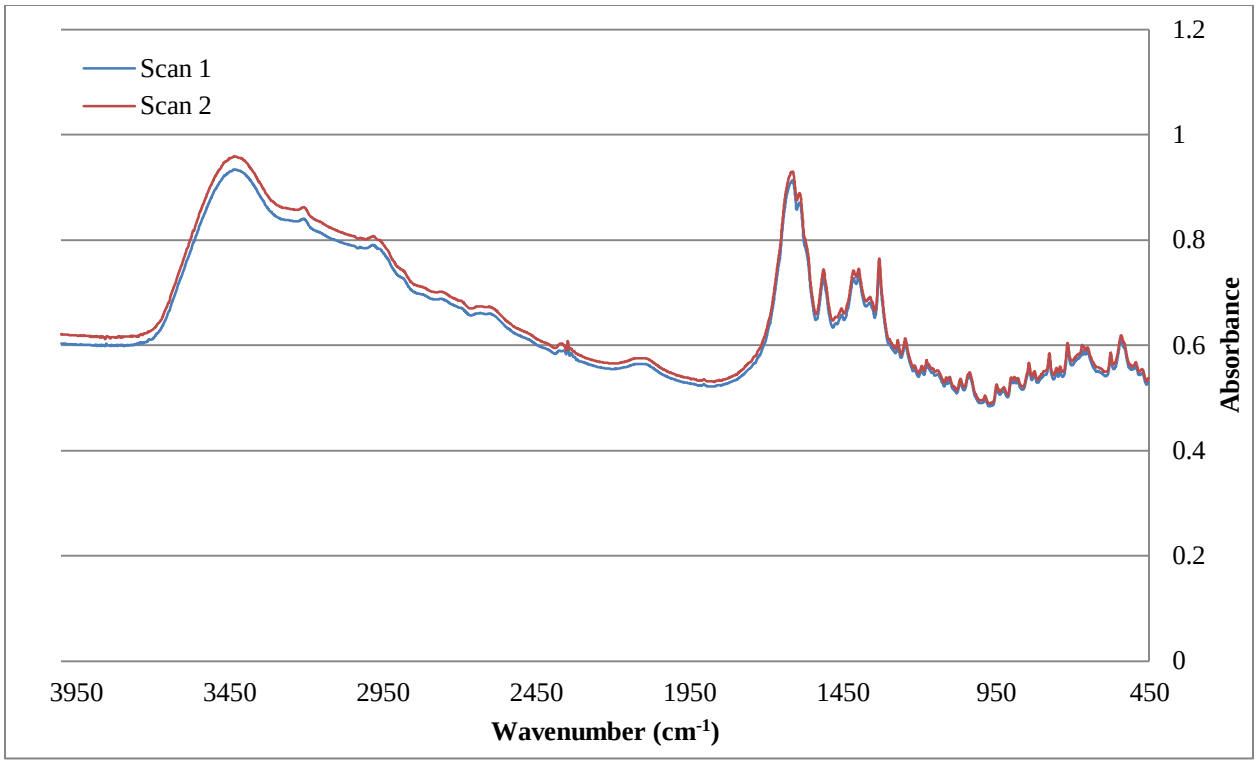


Figure D2. FT-IR spectra of unreacted mixture D. Additional L-tyrosine mixtures were intended to be created, but there was not a sufficient amount of L-tyrosine available to create these mixtures.

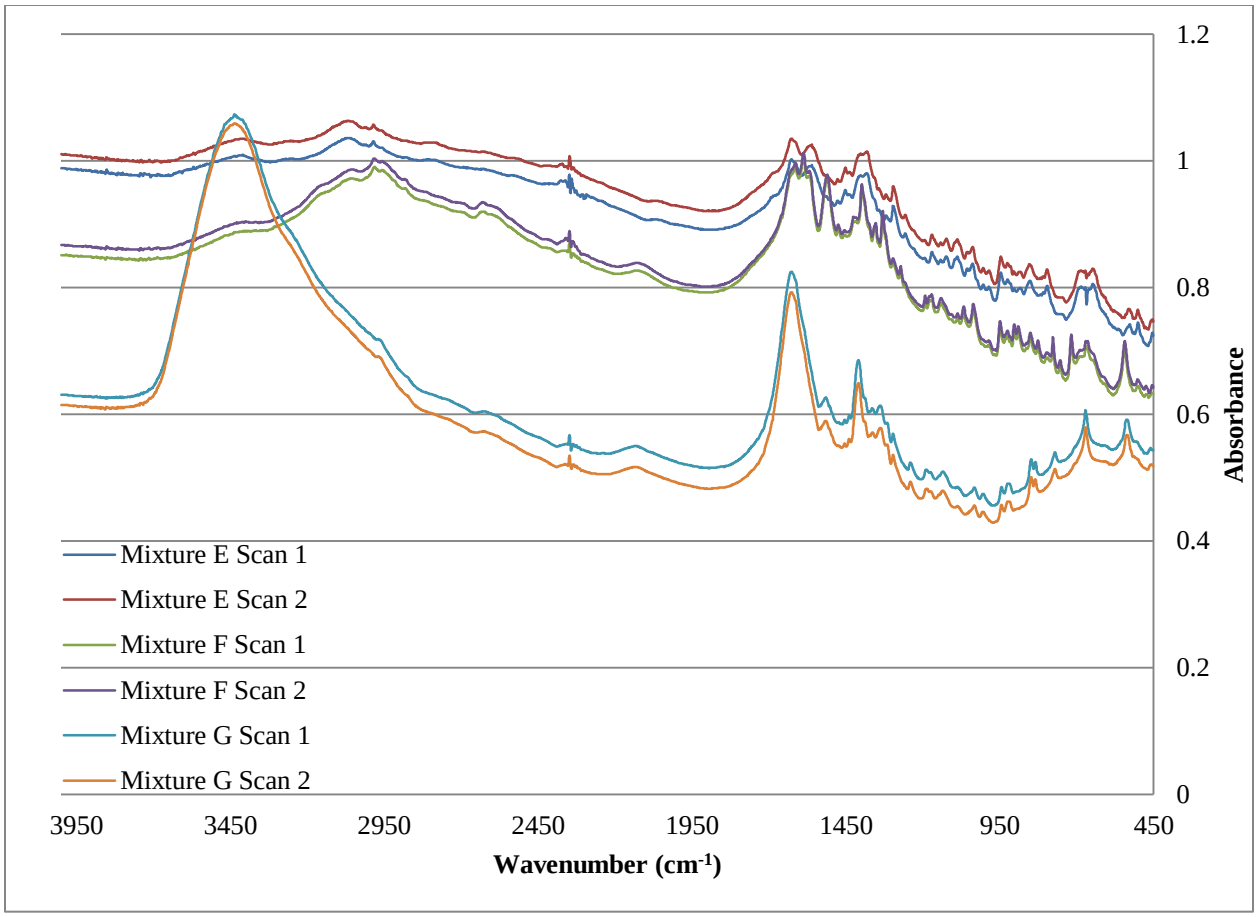


Figure D3. FT-IR spectra of unreacted mixtures E, F and G.

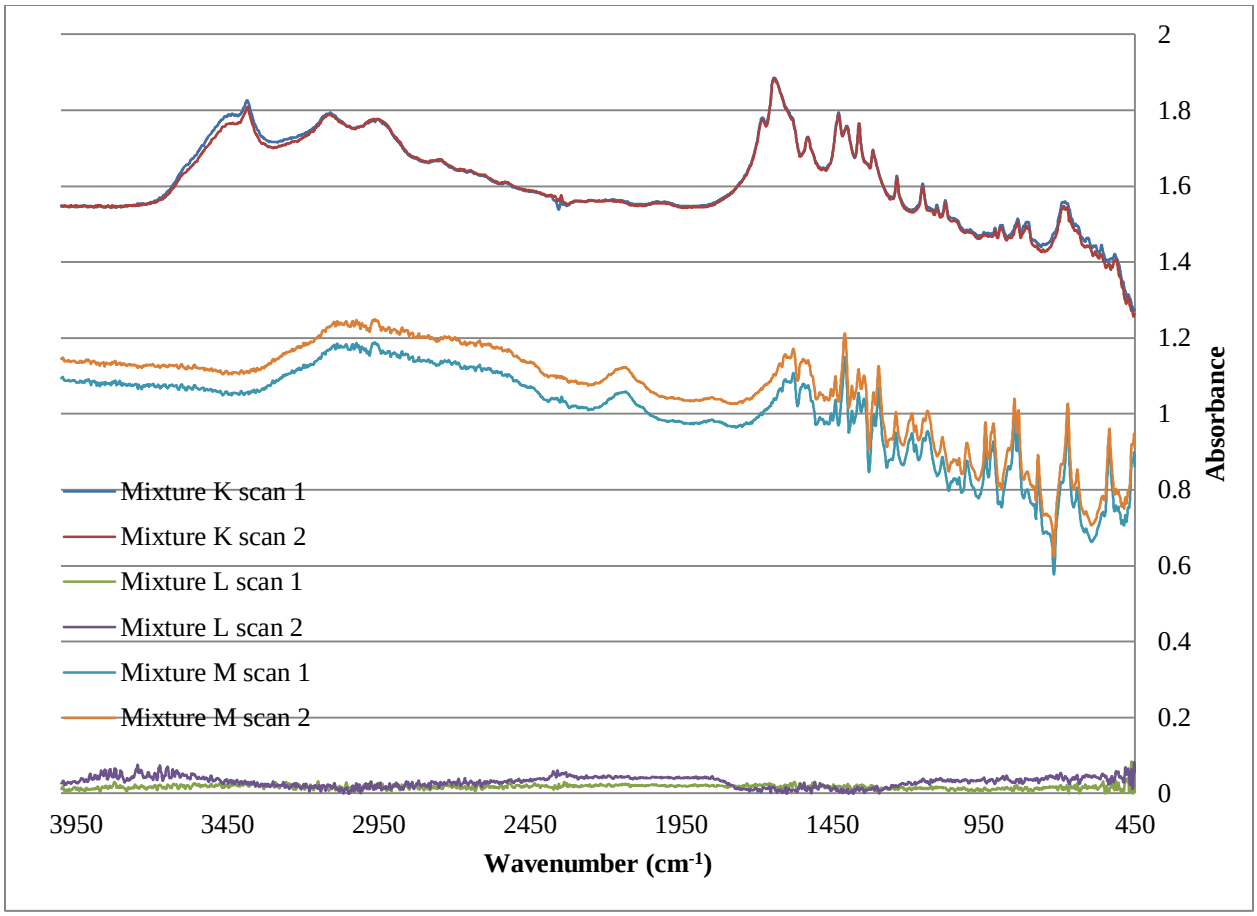


Figure D4. FT-IR spectra of unreacted mixtures K, L and M.

Appendix E. Chemical Algorithm Results of Glycine Reference

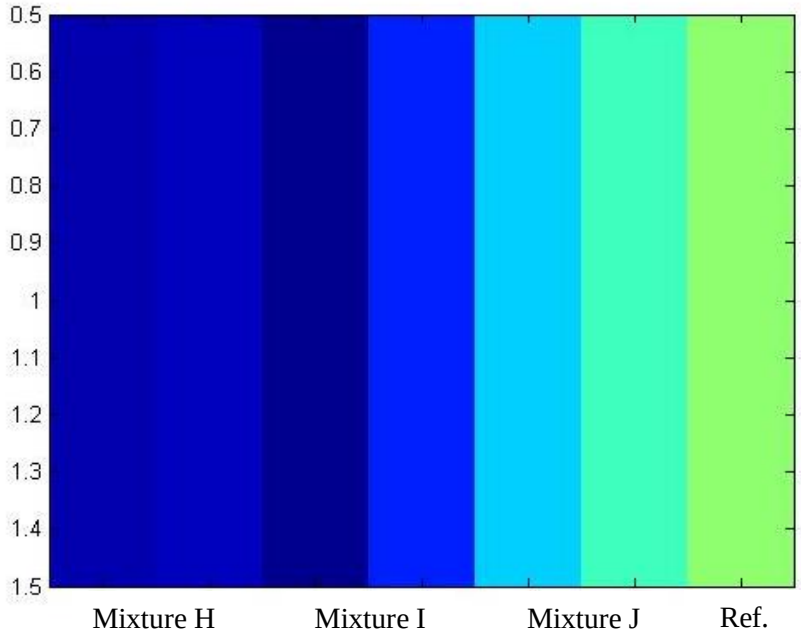


Figure E1. False-color image generated by the algorithm of unreacted mixtures H, I and J spectra: glycine reference.

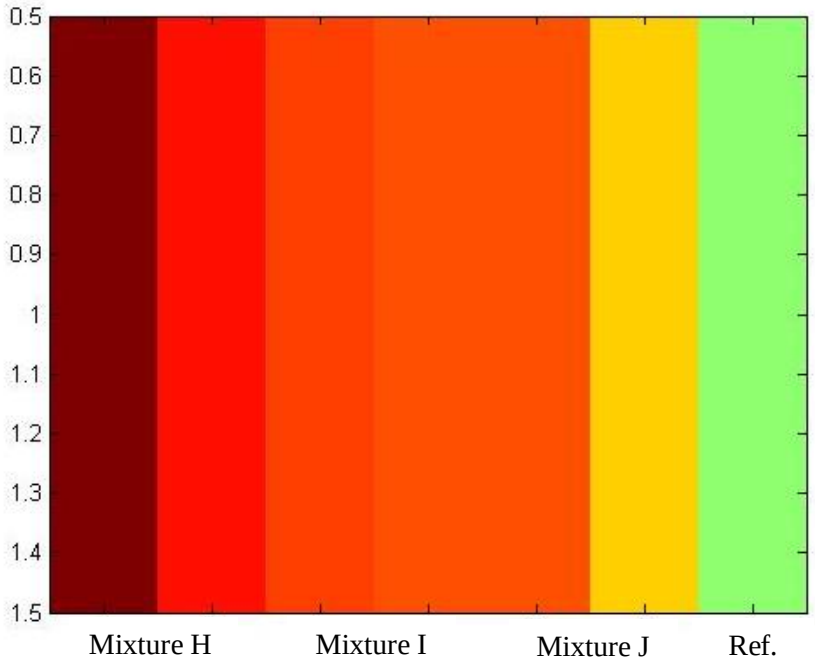


Figure E2. False-color image generated by the algorithm of reacted mixtures H, I and J selected spectra: glycine reference.

Works Cited

1. Abdominal Aortic Aneurysm. https://www.ecu.edu/cs-dhs/surgery/abAneurysm.cfm (accessed 26 Nov 2012).

2. Sakalihasan, N.; Limet, R.; Defawe, O. D., Abdominal Aortic Aneurysm. *The Lancet* **2005,** *365*, 1577-89.

3. Rakita, D.; Newatia, A.; Hines, J.; Siegel, D.; Friedman, B., Spectrum of CT Findings in Rupture and Impending Rupture of Abdominal Aortic Aneurysms. *RadioGraphics* **2007,** *27*, 497-507.

4. What Is Abdominal Aortic Aneurysm? http://www.aorticstents.com/what-is-abdominal-aortic-aneurysm.htm (accessed 19 Jan 2013).

5. Golledge, J.; Muller, J.; Daugherty, A.; Norman, P., Abdominal Aortic Aneurysms: Pathophysiological Mechanisms and Clinical Implications. *Arteriosclerosis, Thrombosis, and Vascular Biology* **2006,** *26*, 2605-2613.

6. Stent-Graft for Repair of Abdominal Aortic Aneurysm. *Interventional Radiology Grand Rounds* **2004**.

7. Yan, W.-D.; Perk, M.; Chagpar, A.; Stratoff, S.; Schneider, W. J.; Jugdutt, B. I.; Tulip, J.; Lucas, A., Laser-Induced Fluorescence: III. Quantitative Analysis of Atherosclerotic Plaque Content. *Lasers in Surgery and Medicine* **1995,** *16* (2), 164-178.

8. Stella, A.; Garguilo, M.; Pasquinelli, G.; Preda, P.; Faggioli, G. L.; Cenacchi, G.; D'Addato, M., The Cellular Component in the Parietal Infiltrate of Inflammatory Abdominal Aortic Aneurysms (IAAA). *European Journal of Vascular Surgery* **1991,** *5*, 65-70.

9. Harris, J. C. New Bioinformatic Techniques for the Analysis of Large Datatsets. Dissertation, University of Kentucky, Lexington, KY, 2007.

10. Gross, D. R., *Animal Models in Cardiovascular Research*. Second ed.; kluwer Academic Publishers: Dordecht, The Netherlands, 1994; Vol. 153.

11. Runkle, R. C.; Tardiff, M. F.; Anderson, K. K.; Carlson, D. K.; Smith, L. E., Analysis of Spectroscopic Radiation Portal Monitor Data Using Principal Component Analysis. *IEEE Transactions on Nuclear Science* **2006,** *53* (3), 1418-1423.

12. Wold, S.; Esbensen, K.; Geladi, P., Principal Component Analysis. *Chemometrics and Intelligent Laboratory Systems* **1987,** *2*, 37-52.

13. Oro, J.; Guidry, C. L., Direct Synthesis of Polypeptides I. Polycondensation of Glycine in Aqueous Solutions. *Archives of Biochemistry and Biophysics* **1961,** *93*, 166-171.

14. Donachy, J. E.; Sikes, C. S., Thermal Polycondensation Synthesis of Biomimetic Serine-Containing Derivatives of Polyaspartate: Potent Inhibitors of Calcium Carbonate and Phosphate Crystallization. *Journal of Polymer Science: Part A: Polymer Chemistry* **1994,** *32*, 789-795.

15. Damm, M.; Glasnov, T. N.; Kappe, C. O., Translating High-Temperature Microwave Chemistry to Scalable Continuous Flow Processes. *Organic Process Research & Development* **2010,** *14*, 215-224.

16. Fox, S. W.; Harada, K., Thermal Copolymerization of Amino Acids in the Presence of Phosphoric Acid. *Archives of Biochemistry and Biophysics* **1960,** *86*, 281-285.

17. Osborne, J. W.; Costello, A. B., Sample Size and Subject to Item Ratio In Principal Components Analysis. *Practical Assessment, Research & Evaluation* **2004,** *9*.